LA-UR-13-22891

Approved for public release; distribution is unlimited.

Title: A Systematic Evaluation of Traditional and Non-Traditional Data Streams for Integrated Global Biosurveillance – Final Report

Author(s):

Deshpande, Alina
Brown, Mac G.
Castro, Lauren A.
Daniel, William B.
Generous, Eric N.
Hengartner, Andrea
Longo, Joseph Francis
Margevicius, Kristen J.
McCabe, Kirsten J.
Parliman, Zachary A.

Intended for: Delivery to Sponsor
Report

Issued: 2013-04-23

A Systematic Evaluation of Traditional and Non-Traditional Data Streams for Integrated Global Biosurveillance – Final Report

Dr. Mac Brown, Ms. Lauren Castro, Dr. William Brent Daniel, Mr. Eric Nicholas Generous, Ms. Andrea Hengartner, Mr. Joseph Longo, Dr. Kristen Margevicius, Dr. Kirsten Taylor-McCabe, Mr. Zach Parliman, and Dr. Alina Deshpande

April 26th, 2013

Contact:
Dr. Alina Deshpande
D-3, Systems Analysis and Surveillance
Defense Systems and Analysis Division
Los Alamos National Laboratory
Mail Stop C936
Los Alamos, NM 87545
Telephone – 505-665-9143/BB: 505-500-2166
E.mail – deshpande_a@lanl.gov

Table of Contents

1.0 Executive Summary

Living in a closely connected and highly mobile world presents many new mechanisms for rapid disease spread and in recent years, global disease surveillance has become a high priority. In addition, much like the contribution of non-traditional medicine to curing diseases, non-traditional data streams are being considered of value in disease surveillance. Los Alamos National Laboratory (LANL) was tasked by the Defense Threat Reduction Agency (DTRA)-Joint Science and Technology Office (JSTO) to determine the relevance of data streams for an integrated global biosurveillance system through the use of defined metrics and methodologies. The long-term, broad objective of this project was to provide information and analysis that can be leveraged from existing and developing national and international disease surveillance systems and methodologies, to create a global disease monitoring system, ultimately providing decision makers with timely information to prepare for and mitigate the spread of disease.

Specifically, this project evaluated data streams that are currently being used in surveillance systems and data streams that had the potential for being used in surveillance to enable early disease detection. LANL's focus was on infectious disease surveillance and the term "biosurveillance" will be used to refer to this scope hereafter. Data stream evaluation was conducted using two different methods: Multi criteria decision analysis (MCDA) using a tool called Logical Decisions® that assigns utility scores to data streams based on weighted metrics and assigned values specific to data stream categories; and a Surveillance Window concept developed at LANL that assigns a window or windows of time specific to a disease within which information coming from various data streams can be determined to have utility. A cross method analysis was performed between the surveillance window based evaluation and the MCDA-based evaluation to identify data stream categories that showed high utility for both methods. In addition, algorithms that can *integrate* useful data streams to facilitate early disease detection were also examined. This project provided an understanding of data streams relevant to early warning, early detection and monitoring of disease outbreaks.

The report describes the results of LANL's evaluation of 16 data stream categories using the two methods, and recommendations for next steps. The project evolved significantly over its course, and significant improvements in our approaches were made compared to the initial proposed processes. A key outcome of this effort was that LANL was able to bring together several diverse entities involved in disease surveillance and lay the foundation for new collaborations that straddled military and civilian health surveillance. The robust evaluation framework developed by LANL has generated significant interest and there is interest in adopting this framework for various applications both within disease surveillance as well as other public health initiatives. Finally, LANL developed the Biosurveillance Resource Directory (BRD), a relational database that underwent pilot testing by members of the human, plant and animal disease surveillance community. The BRD is intended to be a global resource to facilitate rapid information access. The deliverables for the end of the current performance period were;

1) Ranked/Prioritized list of data streams evaluated by MCDA and Surveillance Windows
2) An analysis of algorithms for data stream integration
3) A relational database for surveillance systems and data streams (to include electronic surveillance systems that perform analysis on a combination of disparate data streams leading to actionable results for various surveillance goals).

2.0 Introduction

The evaluation of traditional and non-traditional data streams required a significant effort to first identify data streams and metrics, in order to develop an evaluation framework. During the course of the project, as we gathered data for multiple tasks, it became clear that disease surveillance was not just about early disease detection, but rather, had multiple goals and it was necessary to define these goals to facilitate sound evaluation of data streams. Our search for information yielded a diverse array of term descriptions, opinions and goals that were used to derive LANL definitions for data streams, metrics, and methods to assign values to data streams for individual metrics. Rather than assess specific data sources/data streams such as Google news or Twitter, it was decided that the evaluation be performed at a higher level and specific data streams be binned into general categories. Our approach to data stream evaluation is shown in Figure 1. Surveillance goals, data stream categories and metrics were identified through three main approaches;

1) A comprehensive survey of current and planned surveillance systems that cover human, plant and animal diseases and operate locally, nationally or globally.
2) Establishment of a subject matter expert (SME) panel and panel survey to obtain information about data streams, metrics and biosurveillance goals. Every effort was made to include representatives from the human, animal and plant diseases surveillance community, into the SME panel. Detailed information on the SME panel can be found in Section 13, Appendix E.
3) An extensive review of the scientific literature pertaining to the field of biosurveillance.

The three approaches are described in further detail in Section 3.0 titled “Identification of Data Streams, Metrics and Biosurveillance Goals”. The data stream categories were then evaluated using two different methodologies; Multi criteria decision analysis (MCDA) and surveillance windows, and categories that showed utility with both methods were deemed to be the top ranked data streams. Additionally, LANL evaluated data integration algorithms useful for a global disease surveillance system through a review of scientific literature.

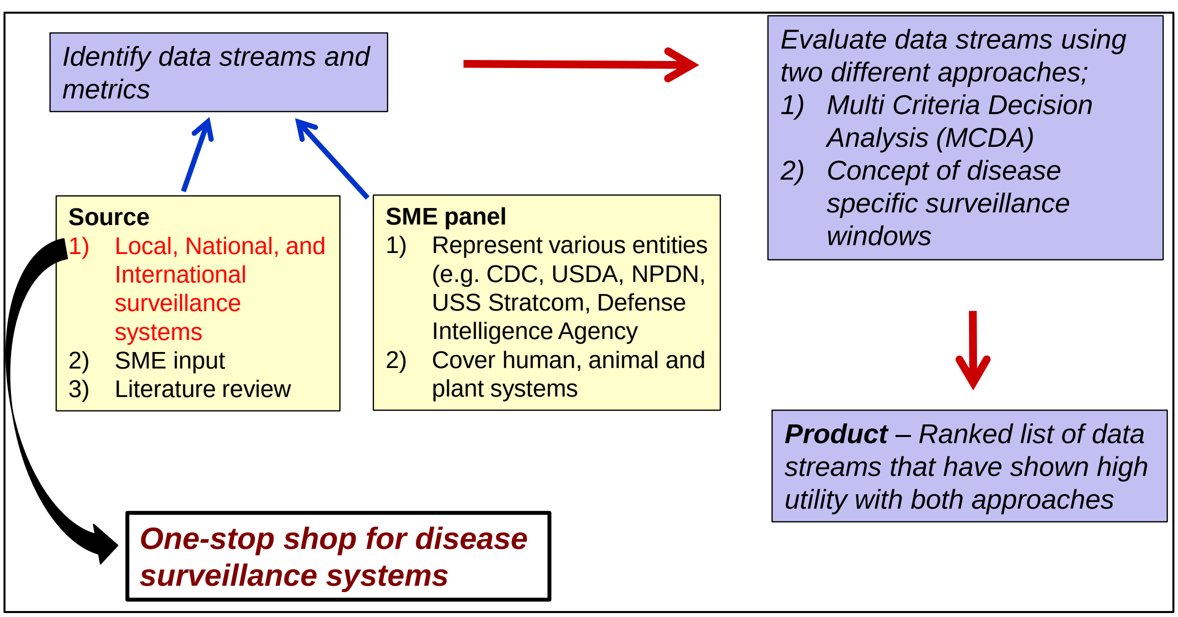


Figure 1: Data stream evaluation approach

A large data collection was produced as a result of the survey of national and international surveillance systems and it covers human, plant, animal and marine surveillance systems. To make this data collection useful to the global biosurveillance community, LANL developed a searchable database that could serve as a "one-stop shop" for disease surveillance resources called the Biosurveillance Resource Directory (BRD). This database could be used by members of the biosurveillance community for validating information they may receive about disease outbreaks. The database is searchable by multiple keywords such as location, data stream use, sponsor, disease, etc. Through consultation with sponsors and other entities involved in biosurveillance, the BRD was deemed to be a very relevant tool and has undergone pilot testing by members of the human, animal and plant biosurveillance community. The overall assessment of the BRD was very favorable and the users would like to have access to the BRD. The BRD is intended to be used by analysts, decision makers (such as base commanders), medical practitioners, public health officials, developers of surveillance systems, and military and civilian members of the biosurveillance community. This database is described in Sections 3.0 and 9.0. A plan has been developed for the continued maintenance, updating, curating and global outreach for the BRD.

We chose to use two completely different methods to evaluate data streams to provide robustness to our results. Data streams showing utility using both MCDA and the surveillance windows would raise to the top of our rankings at the end of the evaluation. Logical Decisions®, based on the MCDA concept, is a decision support tool that uses a systematic method of evaluating alternatives (data streams) based on a series of attributes (metrics), and offers the ability to rank data streams in order of utility, taking into account multiple features desired for an ideal data stream. The MCDA based approach for data stream evaluation is capable of using several metrics, and in terms of data input, needs a list of data streams and metrics, weights assigned to the metrics and values assigned to the data streams for each metric.

A surveillance window can be defined as brief window of time when information gathered can be used to assist decision makers in effectively responding to an impending outbreak. Consequently, information that arrives beyond this window has limited value**.** Surveillance windows are defined using several criteria, and are disease- and operations-specific**,** as well as specific to the goal of the surveillance (early warning, early detection, situational awareness, consequence management). The surveillance window based approach uses the single metric of time to detection to evaluate the data streams, and, in terms of data has more requirements. These include a list of data streams, a list of priority diseases, data from outbreaks (case studies) for each disease, simulations of the disease outbreaks for those diseases that do not have known outbreaks or show gaps in outbreak data, and information about specific data streams available for each case study.

The MCDA based evaluation of data streams is described in detail in Section 4.0 titled "Evaluation of Data Streams Using Multi Criteria Decision Analysis". The surveillance windows based evaluation of data streams is described in detail in Section 5.0 titled "Evaluation of Data Streams Using Surveillance Windows". A review of data integration algorithms is presented in Section 6.0 titled "Data stream Integration Algorithms". Finally, Section 7.0 titled "Progress and Next Steps" summarizes the progress made on this project and immediate next steps that could be performed. Appendices A through F provide additional information on methods used for the evaluations.

3.0 Identification of Data Streams, Metrics and Biosurveillance Goals

As described in Section 2.0, we used three different approaches for identifying specific data streams and metrics. The following paragraphs describe the approaches and results obtained. Over the period of collecting information, it became clear that the world of biosurveillance was vast and diverse, and there was no real standardization of terminology associated with goals of surveillance, surveillance systems, data streams or data sources, integration of data or even metrics to evaluate data streams. Through our consultations with SMEs in the area of biosurveillance, and analysis of collected information, we have developed consensus definitions that we hope can be used in this field and will facilitate a unified understanding of this complex field.

3.1 Survey of Surveillance Systems

Local, national, and global disease surveillance systems have been implemented to meet the demands of monitoring, detecting, and reporting disease outbreaks and prevalence. Varying surveillance goals and geographic reach have led to multiple and disparate systems, each using unique combinations of data sources to meet surveillance criteria. In order to assess the utility and effectiveness of different data streams for global disease surveillance, a comprehensive survey of current surveillance systems was undertaken. For the scope of our project, a biosurveillance system for infectious disease was defined by us as: *an electronic surveillance system that combines disparate data streams or uses a single data stream and performs analysis to report actionable results for various surveillance goals.* We created a framework to broadly categorize biosurveillance goals. Based on our analysis (again primarily through consultation and a thorough literature review) four broad biosurveillance goals were identified: early warning of health threats, early detection of health events, situational awareness (or monitoring), and consequence management. LANL's definitions of biosurveillance goals are the following:

Early Warning of Health Threats: Surveillance that enables identification of potential threats including emerging and re-emerging diseases that may be undefined or unexpected.

Early Detection of Health Events: Surveillance that enables identification of disease outbreaks (either natural or intentional in origin), or events that have occurred, before they become significant.

Situational Awareness/monitoring: Surveillance that monitors the location, magnitude, and spread of an outbreak or event once it has occurred.

Consequence Management: Surveillance that assesses impacts and determines response to an outbreak or an event

Baseline Awareness: Information that can inform and facilitate the achievement of the above surveillance goals and can be related to population demographics and health, the natural, political, and social environment, and underlying disease patterns and characteristics.

The goals tend to follow a time-course from early warning to consequence management, although there is certainly overlap in time. Underlying all of the goals is the need to have baseline awareness of disease and environmental determinants. Figure 2 shows the *goals framework* developed by us and the relationship that these goals have with each other. It is important to note that the broad goals have been linked and overlapped to indicate that there is *no absolute cut off on a time scale when any one biosurveillance goal would be deemed irrelevant*. Likewise, Baseline Awareness is a significant requirement to achieve any of the biosurveillance goals identified in the figure.

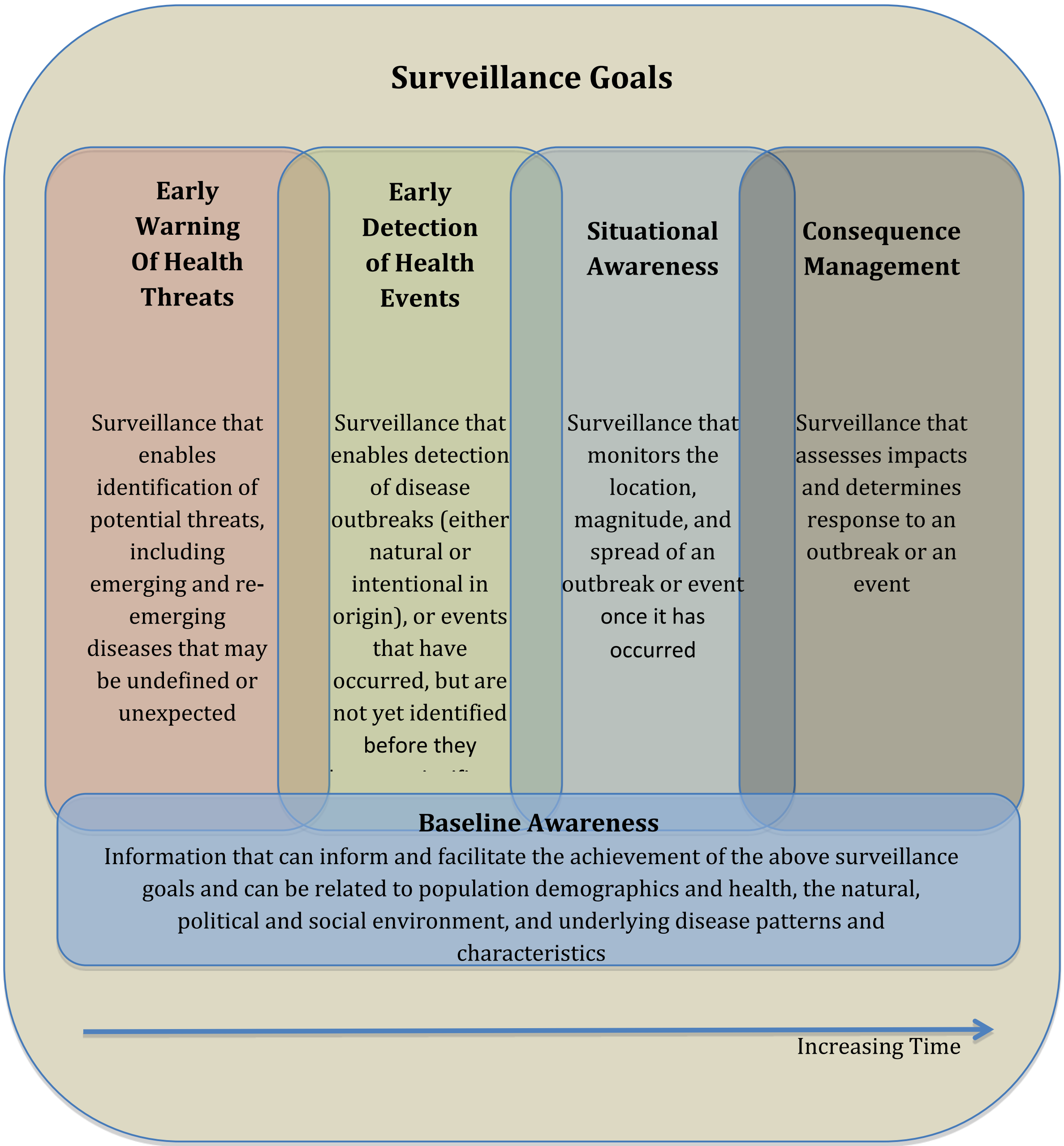


Figure 2: Biosurveillance goals

Initially, we cataloged human, animal, and plant systems in an Excel workbook *(previously delivered*). Over the course of the project this collection has been converted to a relational database (the BRD) built in Filemaker Pro, a commercially available software. To date, 296 items associated with global disease surveillance have been catalogued and include local and national US based surveillance systems, national surveillance systems developed in other countries, as well as surveillance systems that have global coverage. In addition to surveillance systems, we collected other resources that although frequently claimed to be surveillance systems, did not perform analysis of any of the data/information collected and therefore did not fit our definition of an electronic disease surveillance system. We also collected information on tools/algorithms that facilitate analysis of surveillance data. Not surprisingly, the majority of surveillance systems were developed for human diseases. The most frequent sponsoring agencies for these surveillance systems were;

- Centers for Disease Control (CDC)
- Department of Defense (DoD)
- Food and Drug Administration (FDA)
- United States Department of Agriculture (USDA)
- World Health Organization (WHO)
- World Organization for Animal Health (OIE)
- Food and Agriculture Organization of the United Nations (FAO)
- National governments

Table 1 describes the categories of information collected in the BRD, along with examples to illustrate the binning logic. Since our interim report, we have refined the definitions for our information categories as well as the actual categories. The first two categories shown in table 1, "Supersystem" and "System" are the primary types of surveillance systems captured in our database, and served the purpose of providing information on data streams used in disease surveillance. These systems fit our definition as they collect information from one or more systems or data streams analyze the data and inform a biosurveillance goal. During our searches, we came across several "systems" that although collected information from single or multiple data streams, did not perform any analysis and report actionable results and therefore did not fit our definition of a disease surveillance system. Rather than categorizing these entities as surveillance systems, we chose to bin them as a "Data source". The category of information titled "Tool/Software" includes software or applications that facilitate the collection or analysis of data for disease surveillance. Interestingly, a large amount of information searched on the internet under keyword "surveillance system" yielded items that would be categorized as data sources or tools for surveillance. Finally, we collected information about organizations or groups of individuals that contributed to data collection and analysis to inform a biosurveillance goal, and binned them into a category called "Collective".

Information Category	Examples
Supersystem A system that collects information from multiple data streams and other surveillance systems, and analyzes the data that is collected to inform the biosurveillance goal	GOARN TESSy SAGES
System A system that collects information from one or more data streams and analyzes the data that is collected to inform the biosurveillance goal	Biosentinel ASPREN ProMed Health Map
Data Source A system that collects information from one or more data streams but does not analyze the data collected for a biosurveillance goal	Google News Gene Expression Omnibus Crisis Mappers
Tool / Software Software or application that enables the collection or analysis of data	Essence EARS First Watch
Collective A group of individuals or organizations with the shared objective of contributing to data collection and analysis to inform a biosurveillance goal	Mekong Basin Disease Surveillance Wildlife Data Integration Network

Table 1: Information categories collected in the database

While many categories of information are included in the database, a minimum criteria for inclusion required that the information be relevant to biosurveillance (as determined by the surveillance goals) and be available on the internet (has a website). Also important to note is that the BRD is focused on information pertaining to infectious disease surveillance. Details of the BRD are described in Appendix A. The BRD was built over the past two years and has undergone pilot testing by 14 members of the biosurveillance community. A limitation of this pilot test is that the assessments were conducted by US based experts and there is no global representation. This was due to export control regulations. However, the assessments have allowed us to develop a plan for improvement and maintenance of the BRD. The primary organizations that pilot tested the BRD are identified in Appendix B, along with a summary of results. An electronic survey was sent to the evaluators (Appendix C) and detailed feedback for each of the survey questions is a provided in Appendix D. Experts were asked to evaluate the BRD for both content and functionality.

The main features of the current version of the BRD are shown in Figure 3. The relational database contains biosurveillance products and tools available worldwide and is searchable by multiple keywords (geographic location, disease, sponsoring agency, contacts, data streams, etc., is linked to the network, and enables user to directly access the website for the product/tool of interest. Associated reports, factsheets and journal articles can also be linked to if open access is available. We are also considering the configuration of this tool as an application for mobile platforms such as smartphones and tablets.

Figure 3: A depiction of the BRD and its current and future features

Should further investment be deemed appropriate, we have developed a plan to web host the BRD through LANL's Research library, an institutional resource. The research library hosts several databases and has extensive experience with database design and display. LANL would be responsible for monthly updates of the BRD, to include curation of existing records, adding new records and global outreach. The web hosted BRD would be available to the global biosurveillance community. The web link would be sent to all interested parties and we would conduct an operational evaluation of the database to include more members of the global community.

A lesson learned from the pilot testing of the BRD was that as there was no value proposition made to the testers, a significant amount of time was spent in contacting the testers and getting their feedback. The testers did not feel the need to be prompt in their evaluation and LANL had to remind many of them at least three times if not more to send their feedback, making for a very inefficient and time consuming process. Future efforts at getting feedback need to include a monetary compensation or acknowledgement of authorship on a manuscript as a value proposition.

3.2 Survey of Subject Matter Experts (SME panel)

A complete description of our efforts to build an SME panel and the results of the survey are presented in a report in Appendix E. Much like the data collection on surveillance systems, the primary goal of the SME survey was to identify data streams to evaluate and metrics to use for the evaluation. However, we were able to obtain valuable information that facilitated the identification of surveillance goals and definitions of several terms used in the field of biosurveillance. A second electronic survey was sent to biosurveillance experts to obtain their priorities for metrics based on the four biosurveillance goals outlined above. This is described in the Section 4.0.

3.3 Literature Review

Our evaluation of traditional and non-traditional data streams for integrated global biosurveillance required review of the literature in several areas:

- biosurveillance systems, descriptions and evaluations
- data streams in use and considered for use in biosurveillance systems
- diseases, categories of importance in biosurveillance

Category	Number of references (journal articles, conference proceedings, book sections, and government reports)
Biosurveillance, General	376
Biosurveillance, Systems	107
Biosurveillance, Data Streams	44
Diseases, Biosurveillance	115
Diseases, Biosurveillance, Emerging Infectious	111
Diseases, Biosurveillance, Epidemiology	69

Table 2: Summary of literature review

We reviewed over 800 publications that included peer reviewed literature, conference proceedings and book sections, government reports and documents, media reports (newspapers, press releases), and web pages associated with surveillance systems or their associated organizations. A list of categories and number of references associated with each category are listed in Table 2. The references have been compiled as a separate collection. Among those, references considered to be particularly useful for developing data stream categories and metrics were;

1. Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks. Buehler JW., Hopkins RS., Overhage JM., Sosin DM., Tong V. *MMWR*. 53(5):1–11, 2004.

2. Assessing the continuum of event-based biosurveillance through an operational lens. Corley CD, Lancaster MJ, Brigantic RT, et al. *Biosecur Bioterror*. 2012.
3. Animal Health Surveillance Terminology, Draft Output From Pre-ICAHS Workshop, 2011
4. Final Recommendation: Core Processes and EHR Requirements for Public Health Syndromic Surveillance, International Society for Disease Surveillance, 2011

Table 3 shows examples of data streams that were compiled from the literature. They were most commonly referred to as "data sources" and covered diverse types of data.

Population	Type	Data Source	Description	Reference
Human	Environmental /Social -Natural	Internet News Reporting	Real time news event extraction with limited linguistic sophistication for violent and disaster events	Tanev 2008
All	Environmental /Natural	Established Database	Solar radiation, dew point, temperature	Charland 2009, Degroote 2008
Human	Diagnostic	Mobile Lab Report	Mobile phone - geocoded photographs of blood samples for rapid detection of malaria strains	Fuller 2010
Human, Animal, Pathogen	Diagnostic	Lab Report	Biotracing biological contamination in feed/food chain	Hoorfar 2011, Knutsson 2011
Human	Diagnostic	ED/Hospital Report	VA Outpatient Final Diagnosis	Tokars 2010
Human	Syndromic	Internet Search Queries	Search engine queries (ILI)	Ginsberg 2009
Human	Syndromic	Established Monitoring System Internet Search Queries	Google flu trends -internet searches	Ortiz 2011

Table 3: Examples of data streams obtained from literature reviews

3.4 Final Data Streams and Definitions

Multiple data sources are used in a variety of biosurveillance systems that extend from a singular goal for local surveillance of a specific disease to wide-ranging surveillance systems requiring extensive data resources to meet diverse surveillance goals for an array of diseases. With the advent of new technologies, globalization, and high performance computing, there are seemingly unlimited potential data streams that could be useful in biosurveillance. Data streams have not been *universally* defined in either the literature or by specific systems. In order to develop an explicit reference system for data streams, multiple sources of information were gathered, categorized, and analyzed for best use in both describing the data streams and for database development. These included data streams that have been used in biosurveillance systems as described in peer-reviewed literature, as found by our survey of surveillance systems, and as considered important by SMEs. We developed 16 broad categories of data streams to evaluate for this project.

While the sheer number of data streams cataloged was extensive, a framework that would enable each individual data stream to be categorized was developed (Figure 4). The intent was to be able to broadly categorize each individual data stream, but recognize that the data stream does not stand alone, and is part of a larger context. The following terms and definitions were developed by us to facilitate a consistent approach in our evaluation and an invariant frame of reference;

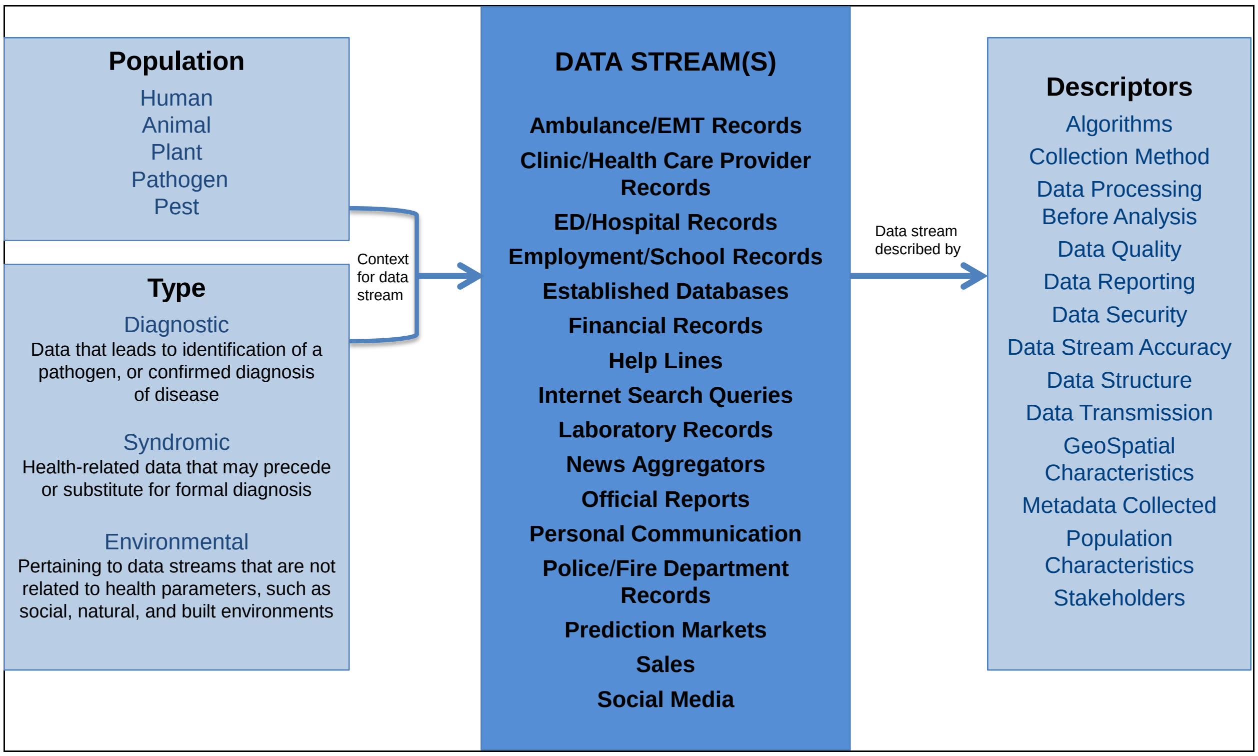


Figure 4: Data stream framework

Data Stream – LANL defines a data stream as a specific source of information that is understood in the context of data population (human, animal, plant, pathogen, and vector) and data type (environmental, syndromic, diagnostic) and is described by specific details such as data structure and collection method.

Population - Human, animal (wild or domestic), plant, pathogen, vector

Type - Determined by the population that the data is coming from and how the data is classified as environmental, syndromic or diagnostic.

a) **Syndromic** -"Health-related data that may precede or substitute for formal diagnosis" (ICAHS, 2011)
b) **Diagnostic** - Data that leads to identification of a pathogen, or confirmed diagnosis of disease
c) **Environmental** - Pertaining to data streams that are not related to health parameters such as *social, built and natural* environments

Social Environment - The social environment, or social conditions in which people live and work, has a major influence on their health. It includes factors such as living conditions, diet, education, and work (WHO http://www.who.int/topics/social_environment/en/). Indicators of the social environment include population demographics, population movements, and political and social engagement.

Built Environment- Distinct from the natural environment, the built environment is comprised of manmade components of people's surroundings, from small-scale settings (e.g., offices, houses, hospitals, shopping malls, and schools) to large-scale settings (e.g., neighborhoods, communities, and cities), as well as roads, sidewalks, green spaces, and connecting transit systems. (Younger et al., 2008)

Natural Environment - the environment that includes vegetation, microorganisms, soil, rocks, atmosphere and natural phenomena as well as universal natural resources and physical phenomena that lack clear-cut boundaries, such as air, water, and climate, as well as energy, radiation, electric charge, and magnetism, not originating from human activity.

Population, Type, and Data Stream categories can be used to characterize the kind of information that is being collected, and how it could impact biosurveillance. Also associated with each individual data stream (if the information is available) are data stream descriptors (see Figure 4). These descriptors are specific to how the data is collected (mobile phones or surveys), how the data is structured, what geographic regions are covered, accessibility and update frequency. All of these descriptors inform the quality and usefulness of the specific data stream. Some examples of data stream descriptors are shown in Figure 5 and examples of data stream collection methods are shown in Figure 6.

Algorithms
Data aggregation
Data trigger/alert

Collection Method

Data Processing Before Analysis
Curated/Uncurated
Collated
Filtered
Transformed
Classified (ICD_9, 10, HL7, Syndromic codes)

Data Quality
Sampling
Consistency
Gaps
Granularity
Completeness
Accuracy
Bias

Data Reporting
On trigger
Irregular
Systematic/ Regular

Data Security
Permissions
Privacy Settings

Data Stream Accuracy
False Positive Rate
False Negative Rate

Data Structure
Structured (relational db)
Semi-structured (XML)
Unstructured

Data Transmission
Lag/no lag time
Raw /aggregated

GeoSpatial Characteristics
Resolution
Scope

Metadata Collected

Population Characteristics

Stakeholders
Source/funding organization
Data providers

Figure 5: Some examples of data stream descriptors

Automated / Electronic Records
Cell / mobile Phone
Data Mining
Database Upload
Email
Field Testing
Internet / web
Manual
Mobile Labs
Phone / Landline
Photographic Images
Remote Sensing
Satellite
Sensors
SMS / Text Messaging
Surveys
Toll Free Phone Access
VHF Radio Relays
Video

Figure 6: Collection methods commonly used in gathering information within each data stream

An advantage of deconstructing data streams in this way is in the ability to discuss data streams in broad terms yet still retain the detailed information that may be very important regarding actual utility of specific data sources. The choice and effectiveness of types of data streams will necessarily be informed by the specific goal(s) of the biosurveillance system, and by the diseases being monitored. Biosurveillance goals and diseases were categorized and defined for the BRD, and their relevance to data stream evaluation is reflected in the data stream framework as shown in Figure 7.

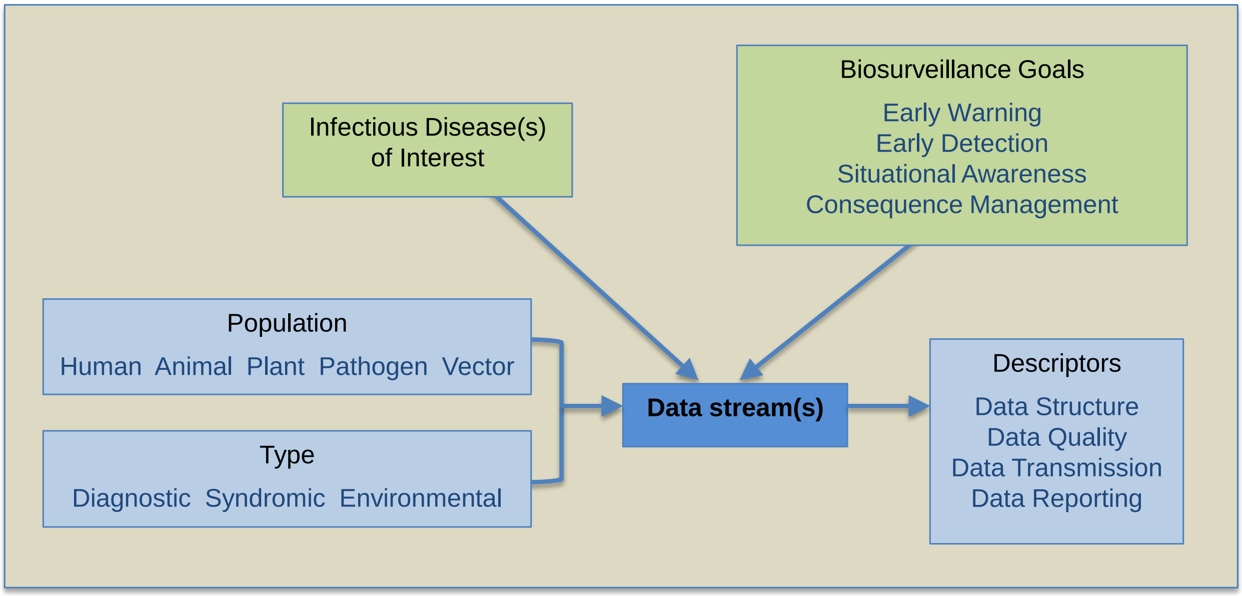


Figure 7: The context and characteristics that describe a data stream.

For example a data stream which monitors Google search queries for health-related key words would be categorized as;

Population: Human

Type: Syndromic

Data stream: Internet search queries

If a data stream was monitoring Twitter for social unrest the stream would be categorized as;

Population: Human

Type: Environmental / Social

Data stream: Social Media/Twitter

Based on our thorough analysis a list of 16 broad categories of data streams were determined (Table 4). We believe that any data source could be binned into one of these categories. Table 4 also shows how the broad categories could be sub-categorized if more information is needed. Specific examples are also given.

DATA STREAM CATEGORY	Sub-Category	Specific Examples
Ambulance / EMT Records Dispatch information which can include incident date, time, nature of call, and patient information		
Clinic/Health Care Provider Records Record of patient (animal/human)	Physician Veterinary	

information that can include symptoms, pharmacy orders, diagnoses, laboratory tests ordered and results received		
ED/Hospital Records Record of patient information that can include discharge/transfer orders, pharmacy orders, radiology results, laboratory results and any other data from ancillary services or provider notes	Military/Veteran Facilities	
Employment/School Records Information collected from schools or places of employment that can include, location, illness, absence, and activity reports regarding students or employees	Absenteeism Illness Activities	
Established Databases Any data repository from which information can be retrieved	Demographic data Geographic data Weather pattern/ meteorlogical data Environmental data Genetic sequencing data	Google Earth Google Maps CIA Factbook Toxnet Census
Financial Records Records of financial activities of a person, business, or organization	Insurance/ HMO billing Bank Savings	
Help Lines Telephone or cellular call-in services	Health/Medical Poison Control Professional Emergency Reporting/Complaint	911 Nurse Hotlines
Internet Search Queries Search terms that a user enters into a web search engine	Global Site Specific	Google Yahoo
Laboratory Records Information regarding specific tests ordered and /or the results of those tests	Laboratory Orders Laboratory Results	PCR Molecular Typing
News Aggregators Systematic collection of information from news sources that can include online and offline media	RSS feeds Radio Video Newspapers Press Releases Media Monitoring	Google News
Official Reports Any report that has been certified or validated from an authorized entity	Government Intelligence Industry Non-profit Academic	WHO CDC / MMWR USDA Notifiable Disease Peer Reviewed Literature
Police/Fire Department Records Dispatch and event information		
Personal Communication Any type of information that is directly relayed from one individual to another individual or group	Expert Non-Expert	

Prediction Markets Marketplaces for contracts in which the payoffs depend on the outcome of a future event	Health Event	Iowa Electronic Health Markets
Sales Monetary transactions for goods or services	Medical Commerical	Drugs (OTC/Rx) Facial Tissue
Social Media Forms of electronic communication such as websites for social networking and blogging through which users create online communities to share	Blogs Internet Chatting Social Networking Sites Video-sharing	Facebook MySpace Twitter YouTube

Table 4: Data Stream Categories

3.5 Final Metrics and Definitions

Similar to our determination of data streams, identifying metrics to evaluate the utility of each data stream was a systematic and iterative process which included a survey of biosurveillance literature relevant to data stream metrics (Bravata et al., 2004; Buehler et al., 2004; Corley et al., 2012; Hitchcock et al., 2007), a survey of data quality analysis literature (Batini et al., 2009; Pipino et al., 2002), an analysis of how metrics/definitions are used in practice from our SME panel, and from the practical application of these metrics in our initial investigation describing and evaluating data streams. Based on the information collected, a description of an *ideal* data stream was developed - A single source of unique, timely (real-time), and spatially relevant information that is standardized and collected in a quantity and class that is needed for meaningful results, that targets a specific population, that is available at many scales (from molecular to ecosystem), is electronically available in both raw and reportable form, and has been rigorously validated. Clearly, no single, specific data stream exists that would have all the features described, but this description facilitated the identification of metrics that we could use in our evaluations.

Table 5 lists the 11 current metrics and definitions that resulted from the above efforts. These metrics and definitions were used and refined through the multi criteria based analysis of the data streams (Section 4.0). Every effort was made to develop metrics that would assess unique features of a data stream and would not overlap. However, it was clear that many of the metrics may have dependencies on each other.

Metric	Description
Accessibility	The extent to which the data stream is available
Cost	The cost to set-up, operate, and maintain the data stream
Credibility	The extent to which the data stream is considered reliable and accurate
Flexibility	The data stream's ability to be used for more than one purpose (such as for use in surveillance for more than one disease, or for more than one goal)
Integrability	How well the data stream can be linked/combined with other data streams
Geographic/Population Coverage	The geographic or population area of coverage
Granularity	The level of detail of the data stream
Specificity of Detection	The ability of the data stream to identify an outbreak, event, disease, or pathogen of interest
Sustainability	The data stream's continued availability over time
Time to Indication	The time required for the data stream to first signal a disease, outbreak, or event
Timeliness	Earliest time that the data is available

Table 5: Metrics to evaluate data streams

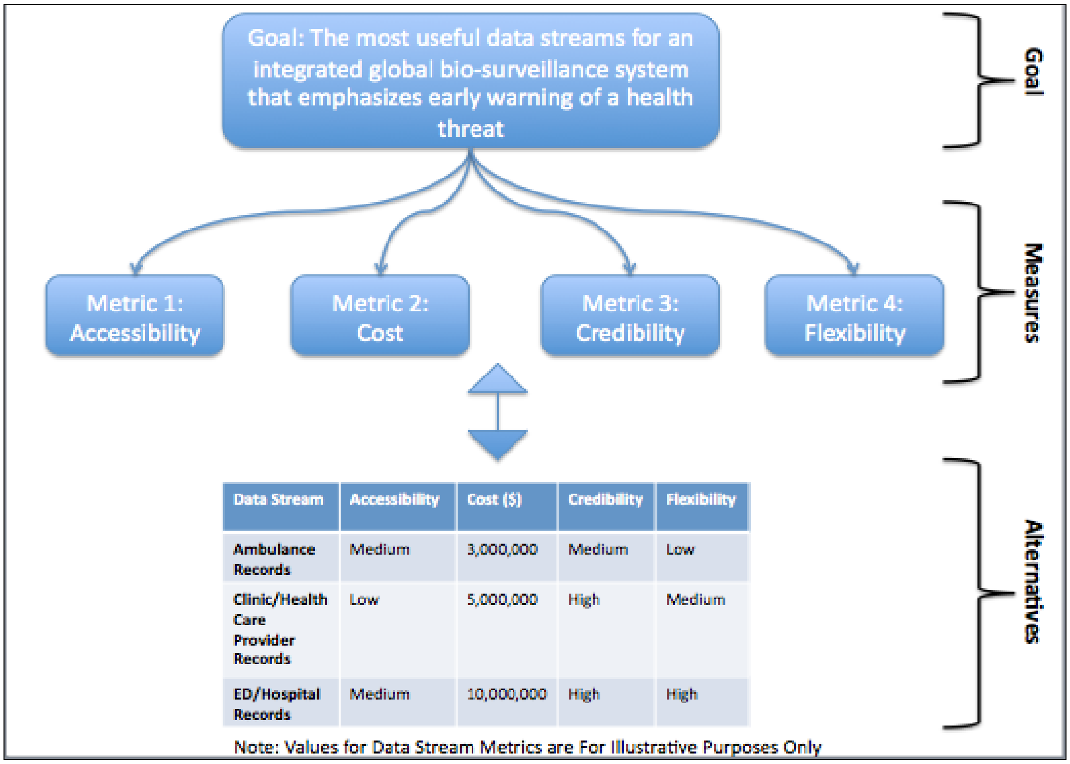
4.0 Evaluation of Data Streams Using Multi Criteria Decision Analysis

Multi-Criteria Decision Analysis (MCDA) allows us to determine the utility and overall desirability of an alternative (options) by the weighted sums of its measures (criteria by which we evaluate the alternatives). This concept is also called Multi attribute utility theory (MAUT) and has been primarily applied in operational decision making. Logical Decisions for Windows® (LDW) is decision analysis software that is based on MCDA and was used for evaluating data streams in this project. By defining alternatives and the measures to describe them it is possible to use MCDA to create a model that can assess and rank the preferences between various alternatives. MCDA/MAUT has been used for applications that range from choosing medical diagnostics (Azar, 2000) to helping retailers create pricing policies for their added value (Wallenius, 2007).

Decision analysis programs are often used for military applications and LDW has been used in particular because of its success rate. Some examples include evaluation by the Air Force of long term mix of technologies, and by the U.S. Army in evaluation of alternatives for destroying stockpiles of toxic gases (http://www.logicaldecisions.com). While use of LDW is commonplace in the field of operations research, it has never been applied to decision making in biosurveillance. However, the application of LDW to evaluate the utility of biosurveillance data streams was deemed to be appropriate as at its core, this evaluation was essentially a multi-objective decision, with the decision being "which data streams would be most useful in an integrated global biosurveillance system". By framing our decision in this context, it was possible to determine a list of metrics and values for those metrics that allowed us to rank the data streams.

4.1 Approach of Evaluation and Results

LDW helps to systematically evaluate, analyze and rank the alternatives/options of interest. In the context of this project, the alternatives were the data streams. To determine which of these is more desirable than others, it was important to define the evaluation measures (i.e. evaluation criteria or metrics) used to calculate a utility score for each data stream. Furthermore, how well each data stream performs for each metric was described. The metrics were further organized into goals to make up a "goal hierarchy" as required by LDW. The goal hierarchy (sometimes also called a value tree, or an objectives hierarchy) is a visual representation of the multi-objective model that is being analyzed. It describes the relationships between the metrics and the goal. Figure 8 depicts this hierarchy which starts with the goal of identifying the most useful data stream for an integrated global disease surveillance system in the top panel and is followed by the identification of the individual metrics that would be used to achieve this goal (cost, accessibility, etc.). The final level in the hierarchy would be the description of the data stream in the context of each metric (e.g. high, low, medium accessibility).



Data Stream	Accessibility	Cost ($)	Credibility	Flexibility
Ambulance Records	Medium	3,000,000	Medium	Low
Clinic/Health Care Provider Records	Low	5,000,000	High	Medium
ED/Hospital Records	Medium	10,000,000	High	High

Figure 8: Diagram illustrating a goal hierarchy

LDW converts the values input for each metric to a common unit termed *utility*. It is important note that the common unit *utility* is not the same as measuring utility (i.e. the "usefulness" of something). *Utility* is the unit that LDW measures and works with in order to determine the overall utility (usefulness) of each alternative (data stream) from the evaluation criteria (metrics). Additionally, the relationship between *utility* and the values input for the criteria need to be defined (a utility function). For example, if the metric is cost, then the *utility* will decrease as the cost increases. The values can be specified as a quantity as well as by labels, which are text descriptions of the possible levels for each metric. Figure 9 depicts the data required to be input for evaluating data streams.

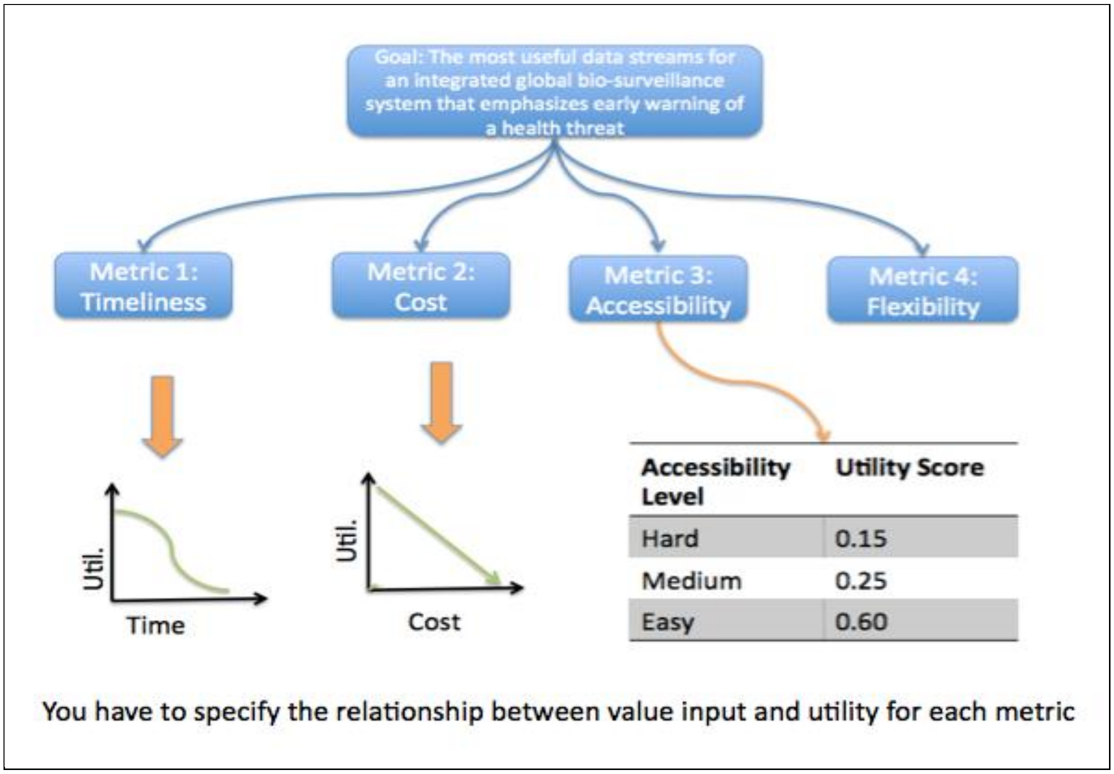


Figure 9: The relationship between metric value and utility needs to be described

In addition to the values input for each metric, the weight of each metric influences the final *utility* score of data streams and also needs to be designated. The different weights of the metrics reflect the varying importance that each individual metric contributes to the final score and, thus, rank of the data streams. Furthermore, the relative weight of each metric can differ depending on context of the biosurveillance goal that the data streams are being evaluated against. For example, the importance of timeliness and time to detection may be higher when the biosurveillance goal is early warning of health threats versus consequence management.

Our approach to the evaluation of data streams followed four broad stages—problem structuring, value elicitation, ranking, and sensitivity analysis—that could be sub-divided into seven steps, each of which were critically important to ensuring high confidence in our rankings;

1. Identify the biosurveillance data streams
2. Identify the biosurveillance goals and objectives
3. Identify the evaluation criteria (metrics)

Problem Structuring

4. Assign the value of each metric for each data stream
5. Assign the weight for each metric

Value Elicitation

6. Rank the data streams
7. Conduct sensitivity analysis

Step 1: Identification of the Biosurveillance Data Streams

As LDW is completely customizable, it was important to scope the question we were trying to answer using this tool. For this project, we were “ranking the usefulness of data streams for an integrated global biosurveillance system”. Based on our definition of data streams (Section 3.4) we identified a multitude of very specific data streams that could each in theory be evaluated using LDW. However, we determined that it would be more prudent to bin the data streams into broad categories or *types* of data streams and evaluate these categories rather than *individual* data streams in order to provide a high value result. In using LDW, our approach would have to balance the level of detail in which we analyzed the data streams between being too high and too low. If we were to analyze each specific data stream , not only would this take an unrealistic amount of time, but the results from such an analysis would be too specific to draw any conclusion about other data streams. Conversely, if we were to approach our analysis too broadly, the results generated would be too vague and not useful. Consequently, our approach attempted to seek an optimal middle ground between these two extremes, which we believe provides the most useful results. Figure 10 is a depiction of our overall strategy for selecting data stream categories/types and the underlying rationale.

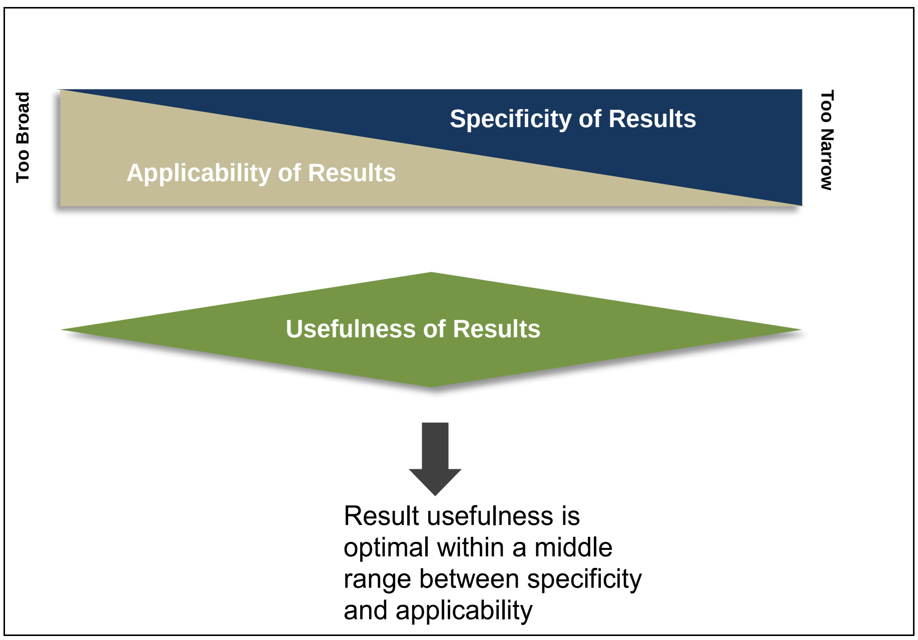


Figure 10: A schematic diagram of the level of detail of our approach

Section 3.4 describes the final 16 categories of data streams we identified. Specific data streams were binned into broader categories of data stream categories, to facilitate evaluation. For example, the category “social media” would contain specific data streams such as Twitter, Facebook, blogs, etc. The category “official reports” would include reports such as ministries of Health, WHO reports, CDC reports, etc.

Step 2: Identification of the Biosurveillance Goals and Objectives

As described in section 3.0, there are multiple goals for disease surveillance and they are arranged over a time scale that extends from pre-event to post-event (the event being a disease outbreak) regardless of origin. It became evident to us that the weights for each of our metrics would be dependent on the biosurveillance goal of interest and it would be necessary to evaluate the data stream types for each surveillance goal as the priorities of metrics such as credibility of the data stream or accessibility may change based on the specific goal being accomplished. The data stream categories were therefore evaluated for each of the following broad surveillance goals, in order of temporal progression;

1. **Early Warning of Health Threats:** Surveillance that enables identification of potential threats including emerging and re-emerging diseases that may be undefined or unexpected.
2. **Early Detection of Health Events:** Surveillance that enables identification of disease outbreaks (either natural or intentional in origin), or events that have occurred, before they become significant.
3. **Situational Awareness:** Surveillance that monitors the location, magnitude, and spread of an outbreak or event once it has occurred.
4. **Consequence Management**: Surveillance that assesses impacts and determines response to an outbreak or an event

Step 3: Identification of the Metrics

Section 3.5 describes a set of 11 metrics that were used to perform LDW based evaluation of the data stream types. We made every effort to ensure that each metric would assess a unique feature of the data stream and were independent from each other.

Step 4: Assigning Values for Each Metric for Each Data Stream Category

Given the highly customizable nature of LDW it was important to scope the problem and be able to obtain a defensible set of rankings for the data stream categories. The concept of "garbage in, garbage out" is equally as applicable to LDW as it is to the field of computer science. This meant that if we were not able to structure the problem correctly and make defensible data input choices, the output of LDW would be meaningless. Assigning specific values to each data stream for a specific metric proved to be an interesting, although not insurmountable challenge as we were evaluating *categories* of data streams not specific ones. How would it be possible to assign specific values to input for each metric in the data streams? Simply averaging the properties of every data stream to derive these values was nearly impossible and impractical. To address this challenge, we decided to focus on the properties of data streams that were functional within commonly used biosurveillance resources collected in the BRD, preferably global ones. The underlying assumption was that the individual, specific data streams within these systems were representative of the data stream *category*. This approach allowed us to derive results that were grounded within the operational context of data streams within current surveillance systems, and while not perfect, allowed us to structure the problem in a way that would yield meaningful results.

A second challenge we faced was how does one assign these values in such a way that the method of measuring them is both constant and meaningful across a variety of data stream categories? (i.e. the metrics and methods of quantifying the data stream's performance equally apply and are meaningful for streams as diverse as social media, ER/ hospital health records, and crowd sourcing, etc.) Thus, for each metric, a method of how to assign the values as well as how those values were related to the common unit *utility* had to be established. Appendix F outlines these methods and relationships as decided by our team.

Representative biosurveillance systems were identified for each data stream category and are listed in Table 6. Some data stream types were better referenced in literature articles.

Data Stream Category	Biosurveillance Resource
Ambulance records	Real-time Outbreak and Disease Surveillance (RODS) System
Clinic/ Health Care Provider Records	ESSENCE
ED/ Hospital Records	Biosense 2.0
Crowd Sourcing	HealthMap
Employment/ School Records	Real-time Outbreak and Disease Surveillance (RODS) System, ESSENCE
Established Databases	Global Pest and Disease Database, World Animal Health Information Database, National Microbial Pathogen Database resource
Financial Records	Real-time Outbreak and Disease Surveillance (RODS) System
Help Lines	FirstWatch
Internet Search Queries	Google Flu
Laboratory Records	ESSENCE
News Aggregators	HealthMap
Official Reports	CDC Reports, Ministry of Health Reports
Police/ Fire Department Records	N/A
Personal Communication	Mekong Basin Disease Surveillance
Prediction Markets	Iowa Health Prediction Market
Sales	National Retail Data Monitor (NRDM)
Social Media	Twitter

Table 6: List of representative resources using various data streams

Based on information obtained from the representative resources for each data stream, along with a review of literature on how the data streams are used, we assigned the following values shown in Table 7. Each member of the LANL team was assigned one or more data streams, and after values had been assigned, a different member of the team checked the values for accuracy. Finally, all team members went through all the data stream category values together, such that the final values were assigned following a consensus. This provided a quality control check for the process.

	Accessibility	Cost	Credibility	Flexibility	Geographic/ Population Coverage	Granularity	Integrability	Specificity of Detection	Sustainability	Time to Indication	Timeliness
Ambulance Records	Medium Accessibility	Medium Cost	Medium Credibility	High Flexibility	Global	Individual	Extremely Integrable	Low Specificity	Yes	Medium Indication	Fast
Clinic/ healthcare Provider Records	Medium Accessibility	Medium Cost	High Credibility	High Flexibility	Global	Individual	Extremely Integrable	High Specificty	Yes	Medium Indication	Fast
ED/ Hospital Records	Medium Accessibility	Medium Cost	High Credibility	High Flexibility	Global	Individual	Extremely Integrable	High Specificty	Yes	Medium Indication	Fast
Employment/ School Records	Medium Accessibility	Medium Cost	Medium Credibility	Low Flexibility	Global	Commmunity	Moderately Integrable	Low Specificity	yes	Indirect Indication	Fast
Established Databases	Easy Accessibility	Low Cost	Low Credibility	High Flexibility	Global	Community	Highly Integrable	Indirect Specificity	Yes	Long Indication	Slow
Financial Records	Medium Accessibility	Medium Cost	Medium Credibility	Medium Flexibility	Regional	Community	Moderately Integrable	Indirect Specificity	Yes	Long Indication	Intermediate
Help Lines	Medium Accessibility	Medium Cost	Medium Credibility	Medium Flexibility	Local	Community	Moderately Integrable	Medium Specificity	Yes	Near Real Time Indication	Fast
Internet Search Queries	Easy Accessibility	Low Cost	Medium Credibility	High Flexibility	Global	Community	Moderately Integrable	Medium Specificity	yes	Near Real Time Indication	Near Real time
Laboratory Records	Medium Accessibility	Medium Cost	High Credibility	Medium Flexibility	Global	Individual	Highly Integrable	High Specificty	yes	Medium Indication	Fast
News Aggregators	Easy Accessibility	Low Cost	Low Credibility	High Flexibility	Global	Community	Moderately Integrable	Low Specificity	yes	Near Real Time Indication	Near Real time
Official Reports	Easy Accessibility	Medium Cost	High Credibility	High Flexibility	Global	Community	Moderately Integrable	High Specificty	Yes	Long Indication	Intermediate
Personal Communication	Easy Accessibility	Medium Cost	Medium Credibility	High Flexibility	Global	Individual	Not Very Integrable	High Specificty	Yes	Long Indication	Fast
Police/ Fire Department records	Difficult Accessibility	Medium Cost	Low Credibility	Low Flexibility	Global	Individual	Moderately Integrable	Indirect Specificity	Yes	Medium Indication	Fast
Prediction Markets	Difficult Accessibility	High Cost	Low Credibility	Low Flexibility	Global	Regional	Moderately Integrable	Medium Specificity	No	Indirect Indication	Fast
Sales	Medium Accessibility	Medium Cost	Low Credibility	High Flexibility	Regional	Community	Moderately Integrable	Low Specificity	Yes	Medium Indication	Fast
Social Media	Easy Accessibility	Low Cost	Low Credibility	High Flexibility	Global	Individual	Moderately Integrable	Low Specificity	Yes	Near Real Time Indication	Near Real time

Table 7: Assigning values for data stream categories

Step 5: Assigning Weights for Each Metric

As mentioned previously, the weights assigned to specific metrics influence the utility score computed by LDW for a data stream, and therefore its ranking. We assigned weights to the 11 metrics following consultation with the SME panel where we asked members to rank metrics in order of importance. In addition, we established an electronic survey and polled LANL biosurveillance experts. They were asked to rank the metrics by the four specific biosurveillance goals we had identifed. Definitions of the metrics and biosurveillance goal were provided. Results were analyzed and a final ranking or priority list for metrics was obtained for each goal, as shown in Table 8.

Early Warning	Early Detection	Situational Awareness	Consequence Management
1. Time to Indication	1. Time to Indication	1. Credibility	1. Credibility
2. Timeliness	2. Timeliness	2. Geo./Pop. Coverage	2. Geo./Pop. Coverage
3. Credibility	3. Credbility	3. Timeliness	2. Timeliness
4. Specificity of Detection	4. Specificity of Detection	4. Time to Indication	4. Specificity of Detection
5. Accessibility	5. Geo./Pop. Coverage	5. Accessibility	4. Time to Indication
6. Geo./Pop. Coverage	6. Accessibility	6. Specificity of Detection	6. Granularity
7. Flexibility	7. Granularity	7. Sustainability	7. Accessibility
7. Granularity	8. Integrability	8. Flexibility	8. Flexibility
9. Integrability	9. Flexibility	9. Integrability	9. Integrability
10. Sustainability	10. Sustainability	10. Granularity	10. Cost
11. Cost	11. Cost	11. Cost	10. Sustainability

Table 8: Ranking of metrics to assign weights

Several trends emerged following analysis of the ranked list of metrics for each of the biosurveillance goals. For early warning of health threats and early detection of health events, both time to indication and timeliness were ranked as the two most important metrics followed by credibility. This contrasts with situational awareness and consequence management, where credibility and geographic/ population coverage were the top two metrics followed by timeliness. Overall, **credibility, timeliness, and time to indication** appeared consistently in the top 3 ranks regardless of surveillance goal and sustainability, cost, and integrability consistently ranked in the bottom four.

LDW converted these rankings into metric weights using a mathematical technique called swing weighting which is used in Simple Multi-Attribute Rating Technique Extended to Ranking (SMARTER). By knowing the rank of the metrics, setting the value for the sum of weights to be 1 and giving equal weights to metrics if the preference is the same (i.e. if multiple metrics are ranked the same.), the weights can be derived for each metric. Table 9 shows the weights assigned to the metrics by LDW.

	Early Warning of a Health Event	Early Detection of a Health Event	Situational Awareness	Consequence Management
Accessibility	0.079	0.067	0.085	0.059
Cost	0.009	0.008	0.008	0.011
Credibility	0.138	0.138	0.275	0.271
Flexibility	0.043	0.027	0.039	0.041
Integrability	0.03	0.039	0.027	0.025
Geo./Pop. Coverage	0.059	0.085	0.184	0.146
Granularity	0.043	0.052	0.017	0.08
Specificity of Detection	0.104	0.108	0.067	0.105
Sustainability	0.019	0.017	0.052	0.011
Time to Indication	0.288	0.275	0.108	0.105
Timeliness	0.188	0.184	0.138	0.146

Table 9: Weights of metrics obtained in LDW using the SMARTER method, the top four heavily weighted are shaded in gray

<u>Step 6: Rank the Data Streams</u>

Data stream ranking was performed through the development of goal hierarchies described previously. As the utility of the metrics is dependent on the context of the biosurveillance goal, we had to design four hierarchies—one for each biosurveillance goal. While the hierarchies were the same for each, the goal specified was different. Following input of weights for metrics, values for each data stream for each metric and a utility function for these values, LDW generated four ranked lists of data streams, one for each surveillance goal, shown in Table 10.

Early Warning of Health Threats	Early Detection of Health Events	Situational Awareness	Consquence Management
1 Internet Search Queries	1 Internet Search Queries	1 ED/Hospital Records	1 ED/Hospital Records
2 ED/Hospital Records	2 ED/Hospital Records	1 Clinic/Healthcare Provider	1 Clinic/Healthcare Provider
2 Clinic/Healthcare Provider	2 Clinic/Healthcare Provider	2 Laboratory Records	2 Laboratory Records
3 Laboratory Records	3 Laboratory Records	3 Internet Search Queries	3 Internet Search Queries
4 News Aggregators	4 News Aggregators	3 Official Reports	3 Official Reports
5 Help Lines	5 Social Media	4 Personal Communication	4 Personal Communication
6 Social Media	6 Ambulance/EMT Records	5 Ambulance/EMT Records	5 Ambulance/EMT Records
7 Ambulance/EMT Records	7 Help Lines	6 News Aggregators	6 News Aggregators
8 Personal Communication	7 Official Reports	7 Social Media	6 Social Media
9 Official Reports	8 Personal Communication	8 Employment/School Records	7 Help Lines
10 Sales	9 Sales	9 Help Lines	8 Employment/School Records
11 Police/Fire Department Records	10 Police/Fire Department Records	10 Financial Recoirds	9 Sales
12 Employment/School Records	11 Employment/School Records	11 Sales	9 Financial Records
12 Financial Records	12 Financial Records	12 Established Databases	10 Police/Fire Department Records
13 Established Databases	13 Established Databases	12 Police/Fire Department Records	11 Established Databases
14 Prediction Markets	14 Prediction Markets	13 Prediction Markets	12 Prediction Markets

Table 10: Ranking of data stream categories using LDW

In general, higher ranked data streams would be considered the most useful, but as can be seen from the table, it is clear that there is no "one size fits all" approach when it comes to biosurveillance. Different data streams vary in utility, and therefore rank, given different biosurveillance goals. There is no one best data stream. However, it is interesting to note which data streams are consistently highly ranked, which trend towards the bottom as well as which change in value when considering each of the biosurveillance goals. Table 11 displays each data stream and its rank for each of the four goals. Data streams shaded in blue indicate they were consistently across four goals ranked among the top 5, whereas data streams shaded in green indicate streams that were ranked in the top 5 at least one goal.

Data Stream	Early Warning of a Health Threat	Early Detection of a Health Threat	Situational Awareness	Consequence Management
Internet Search Queries	**1**	**1**	**3**	**3**
ED/Hospital Records	**2**	**2**	**1**	**1**
Clinic/Healthcare Provider	**2**	**2**	**1**	**1**
Laboratory Records	**3**	**3**	**2**	**2**
News Aggregators	**4**	**4**	6	6
Help Lines	**5**	7	9	7
Social Media	6	**5**	7	6
Ambulance/EMT Records	7	6	**5**	**5**
Personal Communication	8	8	**4**	**4**
Official Reports	9	7	**3**	**2**
Sales	10	**3**	11	9
Police/Fire Department Records	11	10	12	10
Employment/School Records	12	11	8	8
Financial Records	12	12	10	9
Established Databases	13	13	12	11
Prediction Markets	14	14	13	12

Table 11: A comparison of rankings obtained for each data stream category by surveillance goal

Across the four biosurveillance goals, there was a dichotomy exhibited between data stream category ranks in the early warning/early detection goals and the situational awareness/consequence management goals. As observed in Table 11, the ranks for the data streams are fairly consistent (i.e. the same) within the early warning/early detection goals and within the situational awareness/consequence management goals. This seems to suggest that while we identified four distinct biosurveillance goals, functionally there may only be two: pre- and early event (i.e. the initial stages of an outbreak) and post event. Interestingly, this phenomenon correlates with the surveillance window concept that was used as our second method for data stream evaluation.

Four data stream categories ranked within the top 5 for every single goal: **Internet Search Queries, ED/Hospital Records, Clinic/Healthcare Provider, and Laboratory Records.** Three of these—ED/Hospital Records, Clinic/Healthcare Provider, and Laboratory Records—are commonly used in current systems; only Internet Search Queries are not currently used as a data stream in systems. However, given their novelty it is not entirely surprising and it may take time before this data stream category is adopted as a reliable source for systems.

There were four data streams that ranked consistently among the similar goals (early warning/detection vs. situational awareness/consequence management): **Official Reports, Personal Communication, News Aggregators, and Ambulance/EMT records**. Official Reports were ranked quite high for both situational awareness and consequence management, mostly due to the high values assigned for credibility and specificity of detection. While ranked highly for the situational awareness and consequence management goals, Personal Communication ranked towards the middle for the early warning and early detection goals. However, in our discussions with epidemiologists and biosurveillance practitioners, Personal Communication was often cited as one of the most important data streams they utilized to detect outbreaks in its early stages and to monitor its progress. Personal communications tend to be informal, highly unique and diverse in nature making it difficult to sign attributes using our approach—analysis of categories of data streams. A better understanding of the nature of these informal personal communication networks and what roles they may play in the decision making process leading up to an outbreak declaration may lead to some valuable insights.

Social Media, Help Lines, and Sales data streams were all ranked at least once among the top 5. After these data streams, there was a significant drop off in the ranks. In particular, five data streams were consistently ranked as being the least useful: Financial Records, Established Databases, Prediction Markets, Employment/School Records and Police/Fire Department Records. It is important to note that while certain data streams ranked low, it does not mean they are useless. Certain data streams such as Financial Records and Established Databases may be very useful when used *synergistically* with other, more highly ranked data streams. One limitation to the MCDA approach is that it does not take the synergy of data streams into account but instead treats each data stream as if it were independent.

While the MCDA approach offered a very systematic approach to evaluating data streams and forced a much deeper analysis of metrics than has been considered for biosurveillance, there are limitations to this approach that must be carefully considered and accounted for during the review of these results;

1) The LDW tool is highly dependent on user input to structure the problem and elicit the values, and if the input is made without a defensible reason, the results can be of very little value. Thus, at every step of the process, efforts were made to ensure that the choice made for each decision could be defended. Additionally, when we were determining the values to input for the metrics, we focused on using values and properties of data streams in use within a surveillance

resource that we thought were representative of that type of data stream category. Because of this, the results maybe slightly biased towards more traditional data stream categories.

2) There was a bias in our SME panel, which predominantly consisted of experts in human health, representing the developed world, and was largely academic. As a result, their opinions on metric weights and definitions may not jive completely with the operational world. For example, “specificity of detection” was a metric that was unanimously accepted as a useful one by our SME panel. However, following recent consultations with practitioners and public health officials, we found that the metric was not considered a high priority primarily because they did not understand the definition of the metric and its context. We are currently trying to establish a more diverse and operational SME panel in collaboration with the International Society for Disease Surveillance (ISDS), and would like to conduct a second evaluation of the data streams by taking their opinions onto account. It is possible, that a significant disparity in the results is obtained following a second evaluation, and this will bring to light gaps in our understanding of **operational biosurveillance** and its needs.
3) Each metric is considered independently, and related metrics that have an influence on each other cannot be evaluated jointly. For example, accessibility and cost are likely related if accessibility to a data stream is associated with a subscription or user fee. In such a case, a lower cost would let one assign an “easy” qualitative value to the accessibility metric, and vice versa. Such dependencies cannot be input into the LDW tool. While it is possible to model these relationships using MCDA, it would have been a significant undertaking to elucidate the precise nature of the interdependencies amongst the metrics and was beyond the scope of this project.
4) The MCDA approach does not model synergy of data streams. Like simulating the interdependencies of the metrics, there may be synergistic effects when utilizing multiple, different but complementary data streams. Malaria forecasting is an example of this—knowing that there was heavier than normal rainfall in a malaria prone region would likely make officials declare an outbreak quicker if they notice a slight increase in malaria cases. As a result of this, data streams such as Established Databases that were ranked lowly may be extremely useful, however only if they are combined with another data stream.
5) The use of a representative, currently operational surveillance system was used as a model system to determine the values for the metrics of the data streams. Since we were not ranking specific, individual data streams but actually categories of data streams, this was necessary to evaluating data streams for use in an integrated, global biosurveillance system. However, there are some limitations to this. By using currently operational systems, our results may be biased towards more traditional types of data streams currently employed in the developed world.

<u>Step 7 Conduct Sensitivity Analysis</u>

LDW is a tool that relies heavily on user input and customization and the rankings reported in this study may be influenced by the input parameters. It was important to make sure that the rankings were robust to variations in these parameters, and therefore of high confidence. Sensitivity analysis was conducted by varying the dependent variables to understand their influence on data stream rankings. The following strategies were used for this analysis. It is important to note that all changes for each strategy were applied simultaneously rather than looking at the effect of one variable sequentially, in order to maintain a realistic scope for the number of LDW runs for each strategy ;

1) **Varying the utility function** that describes the relationship between metric value and *utility;* the utility functions for metrics that were weighed higher overall were modified from the default linear function assigned to every metric. By varying the utility function, it is possible to assess the impact of our assumptions on the relationship between the metric value and *utility*. For the rankings produced in step 6, each of the metrics values were described using two to four labels. The correlation between the labels and *utility* score was linear. Thus, if there were two labels, the lower option was assigned a *utility* score of 0 and the higher option was assigned a *utility* score of 1. If there were three labels, the lower option was assigned a *utility* score of 0, the intermediate option was assigned a *utility* score of 0.5, and the highest option assigned a *utility* score of 1, so on and so forth. However, this assumption, that the relationship between label and *utility* score is linear, may not be correct. Perhaps after a certain threshold, the *utility* score levels off. We changed the *utility* function for five of the metrics: **integrability, credibility, specificity of detection, time to indication, and timeliness** that were weighed higher overall and therefore may have the greatest impact on rankings. For example, for timeliness and time to indication, the rationale was that as long the data stream was available or indicated an event within one week, the added benefit by detecting it earlier was marginal. Similarly, in the case of credibility a case could be made that only data streams that did not need validation in order to be actionable, or a highly credible data stream by our definition, was most desirable, whereas any data stream that needed any sort of validation, minimal or not, was not nearly as desirable.

2) **Varying weights of metrics;** changing the weights in two ways assessed the impacts of the metric weights. The first was to set all metric weights equally so that they each metric contributed to the final *utility* score. The second was to group the rankings of the metrics into three tiers (Table 12). For each biosurveillance goal, metrics that were in the top 5 ranks were assigned a single high weight, metrics in the middle ranks were assigned a single medium weight and those that ranked in the bottom 3-4 were assigned a single low weight. Given how the SMARTER method decomposes the ranks into weights, instead of having eleven different ranks, grouping the list into three tiers of metrics might be more accurate.

Goal 1		Goal 2		Goal 3		Goal 4	
0.155	Specificity of Detection	0.161	Specificity of Detection	0.161	GEO/POP	0.145	Specificity of Detection
0.155	Credibility	0.161	Credibility	0.161	Credibility	0.145	GEO/POP
0.155	Time to Indication	0.161	Time to Indication	0.161	Time to Indication	0.145	Credibility
0.155	Timeliness	0.161	Timeliness	0.161	Timeliness	0.145	Time to Indication
0.072	Flexibility	0.078	GEO/POP	0.078	Specificity of Detection	0.145	Timeliness
0.072	GEO/POP	0.078	Granularity	0.078	Accessibility	0.078	Granularity
0.072	Granularity	0.078	Accessibility	0.078	Sustainability	0.078	Accessibility
0.072	Accessibility	0.03	Flexibility	0.03	Flexibility	0.03	Flexibility
0.03	Cost	0.03	Cost	0.03	Cost	0.03	Cost
0.03	Integrability	0.03	Integrability	0.03	Granulairty	0.03	Integrability
0.03	Sustainability	0.03	Sustainability	0.03	Integrability	0.03	Sustainability

Table 12: Weights assigned to three tiers of metrics for each goal

3) **Performing rankings without Geographic/Population metric;** for each data stream with the exception of three, Geographic/Population coverage was uniformly assigned a value of "Global". In order for a metric to be effective, it must be able to distinguish and segregate the options it is describing. This suggested that Geographic/Population coverage may not be a useful evaluation criterion, even though many people thought it was important. To see what impact this metric had on the final rankings, the rankings were recomputed without the Geographic/Population coverage metric.
4) **Changing the most variable metric values in the matrix;** we assigned values to data streams for each metric, using representative biosurveillance resources that routinely used specific data streams. In order to understand the variability in how data streams like Official reports were used at a global level, it was necessary to review more than one resource for each data stream. In doing so, it became clear that there were some data streams that showed high variability in assigned values for certain metrics. To examine the influence of variable values on the final ranking of data stream categories, we first ran LDW with an input of all high values for the data streams that showed most variability in certain metrics and then with an input of all low values.

Tables 13-16 show the comparison of rankings obtained in the original run of LDW with the rankings obtained following sensitivity analysis, for each of the four biosurveillance goals. Overall, with the different sensitivity analyses, the results of the modified rankings suggest that the results obtained in the final rankings in step 6 are robust. The same data streams that tend to be ranked as being most useful remain the top ranked. Similarly, the same data streams that tend to be ranked in the middle and at the bottom in the final rankings are observed to do the same in the modified rankings.

Early Warning of a Health Threat	Final Rankings	Without Geographic/ Population Coverage	Varying the Utility Function	3 Tiers of Metric Weights	Equal Weights	Highest Rank	Lowest Rank
ED/ Hospital Records	**2**	**2**	**1**	**1**	**1**	**1**	**2**
Clinic/ Healthcare Provider	**2**	**2**	**1**	**1**	**1**	**1**	**2**
Laboratory Records	**3**	**4**	**2**	**3**	**3**	**2**	**4**
Internet Search Queries	**1**	**1**	**3**	**2**	**2**	**1**	**3**
Official Reports	9	8	11	7	7	7	11
Personal Communication	8	7	9	**4**	6	**4**	9
Social Media	6	**5**	7	6	**5**	**5**	7
News Aggregators	**4**	**4**	**5**	**5**	**4**	**4**	5
Ambulance/EMT Records	7	6	**4**	9	**5**	**4**	9
Help Lines	**5**	**3**	6	8	8	**3**	8
Sales	10	9	8	10	9	8	10
Employment/ School Records	12	12	12	11	10	10	12
Police / Fire Department Records	11	10	10	12	12	10	12
Financial Records	12	11	15	12	11	11	15
Established Databases	13	13	14	13	8	8	14
Prediction Markets	14	14	13	14	13	13	14

Table 13: Comparison of data stream rankings for Early Warning surveillance goal

Early Detection of a Health Threat	Final Rankings	Without Geographic/ Population Coverage	Varying the Utility Function	3 Tiers of Metric Weights	Equal Weights	Highest Rank	Lowest Rank
ED/ Hospital Records	**2**	**2**	**1**	**1**	**1**	**1**	**2**
Clinic/ Healthcare Provider	**2**	**2**	**1**	**1**	**1**	**1**	**2**
Laboratory Records	**3**	**3**	**2**	**2**	**3**	**2**	**3**
Internet Search Queries	**1**	**1**	**3**	**3**	**2**	**1**	**3**
Official Reports	7	7	11	**4**	7	**4**	11
Personal Communication	8	7	9	**5**	6	**5**	9
Social Media	**5**	**5**	6	7	**5**	**5**	7
News Aggregators	**4**	**4**	**5**	6	**4**	**4**	6
Ambulance/EMT Records	6	6	**4**	8	**5**	**4**	8
Help Lines	7	**5**	7	9	8	**5**	9
Sales	9	8	8	10	9	8	10
Employment/ School Records	11	11	12	11	10	10	12
Police / Fire Department Records	10	9	10	12	12	9	12
Financial Records	12	10	15	13	11	10	15
Established Databases	13	12	14	14	8	8	14
Prediction Markets	14	13	13	15	13	13	15

Table 14: Comparison of data stream rankings for Early Detection surveillance goal

Situational Awareness	Final Rankings	Without Geographic/ Population Coverage	Varying the Utility Function	3 Tiers of Metric Weights	Equal Weights	Highest Rank	Lowest Rank
ED/ Hospital Records	**1**	**1**	**1**	**2**	**1**	**1**	**2**
Clinic/ Healthcare Provider	**1**	**1**	**1**	**2**	**1**	**1**	**2**
Laboratory Records	**2**	**2**	**2**	**3**	**3**	**2**	**3**
Internet Search Queries	**3**	**2**	**3**	**1**	**2**	**1**	**3**
Official Reports	**3**	**3**	**4**	**5**	7	**3**	7
Personal Communication	**4**	**4**	6	7	6	**4**	7
Social Media	7	8	8	6	**5**	**5**	8
News Aggregators	6	7	7	**4**	**4**	**4**	7
Ambulance/EMT Records	**5**	6	**5**	7	**5**	**5**	7
Help Lines	9	**5**	11	8	8	**5**	11
Sales	11	9	9	9	9	9	11
Employment/ School Records	8	9	9	9	10	8	10
Police / Fire Department Records	12	11	9	10	12	9	12
Financial Records	10	10	13	11	11	10	13
Established Databases	12	12	10	12	8	8	12
Prediction Markets	13	13	12	13	13	12	13

Table 15: Comparison of data stream rankings for Situational Awareness surveillance goal

Consequence Management	Final Rankings	Without Geographic/ Population Coverage	Varying the Utility Function	3 Tiers of Metric Weights	Equal Weights	Highest Rank	Lowest Rank
ED/ Hospital Records	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Clinic/ Healthcare Provider	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Laboratory Records	**2**	**2**	**2**	**3**	**3**	**2**	**3**
Internet Search Queries	**3**	**3**	**4**	**2**	**2**	**2**	**4**
Official Reports	**3**	**4**	**3**	**4**	7	**3**	7
Personal Communication	**4**	**5**	**5**	**5**	6	**4**	6
Social Media	6	8	8	7	**5**	**5**	8
News Aggregators	6	8	7	6	**4**	**4**	8
Ambulance/EMT Records	**5**	7	6	8	**5**	**5**	8
Help Lines	7	6	9	9	8	6	9
Sales	9	10	10	10	9	9	10
Employment/ School Records	8	9	12	10	10	8	12
Police / Fire Department Records	10	12	11	11	12	10	12
Financial Records	9	11	15	12	11	9	15
Established Databases	11	13	14	12	8	8	14
Prediction Markets	12	14	13	13	13	12	14

Table 16: Comparison of data stream rankings for Consequence Management goal

4.2 Next Steps

A significant part of the MCDA-based evaluation of data streams was devoted to defining clear metrics to use for ranking the data streams, as well as collecting defensible data to support assigning values for each of the metrics. In addition, prioritization of the metrics was helped by defining clear biosurveillance goals. We ranked data stream categories and tested the robustness of these rankings through sensitivity analysis as described in step 7. To provide a “real world” assessment of the data stream categories, a proposed **immediate next step** for this task will be to form a panel of operational biosurveillance experts through our collaboration with the International Society for Disease Surveillance (ISDS) and survey them for metric weights and data stream values. We propose to include representatives of the global human, plant and animal disease surveillance communities to also provide a more “global” context. We will then compare rankings for data streams obtained with the new set of data, with the current rankings. In addition, we will evaluate specific data streams deemed to be priority by DTRA and the Armed Forces Health Surveillance Center (AFHSC). These two tasks will complete the evaluation process using the MCDA based method.

The MCDA-based evaluation method and our framework have generated significant interest not only in the biosurveillance community, but also in the general public health community. As a result there are three entities who are interested in adopting LANL’s evaluation framework for their projects;

1) The ISDS would like to help refine our framework and adopt it for evaluating both specific novel data streams that come on line for state and local disease surveillance and existing data streams currently in use.

2) Dr. Courtney Corley from Pacific Northwest National Laboratory (PNNL) is tasked with evaluating disease forecasting and prediction models by the Department of Homeland Security (DHS), for a specific number of priority diseases. He has approached LANL and obtained the set of metrics used by us for data stream evaluation. With LANL's help, PNNL will be modifying these metrics to use for their evaluation task
3) Dr. Peggy Honore from the Department of Health and Human Services has requested information on our evaluation framework to assist in evaluating US State public Health Systems

5.0 Evaluation of Data Streams Using Surveillance Windows

In the second approach to data stream evaluation, we used a new method of evaluating the utility of data stream categories – the concept of the "surveillance window", defined as the brief period of time when information gathered can be used to assist decision makers in effectively responding to an impending outbreak. The single metric used in this approach was "time to indication" or how early is the data from a particular data stream category available for identifying an event of concern.

Figure 11 shows the overall approach to using this method for evaluating data stream categories. Our first step was identifying a list of priority diseases to select case studies and build surveillance windows, and our primary sources were our SME panel as well as CDC and DOD priorities. Evaluating data streams across multiple diseases would provide a robust application of our method as well as allow for the examination of the influence of disparate disease characteristics not only within human, but across species, on outbreak progression and surveillance window duration. We also conducted a literature review to support our selection of diseases. We ensured that there was representation of human, animal and plant diseases and there was enough data available for selected outbreaks/case studies to facilitate evaluation of the maximum number of data stream categories available for that case study. Below is a list of diseases that were selected;

- Influenza
- Dengue
- Ebola
- Cholera
- Food poisoning due to *E.coli*
- Lassa Fever
- Foot and Mouth Disease (FMD)
- Citrus Greening
- Malaria
- West Nile Virus

Case studies were selected for each disease by reviewing historical outbreaks around the world. An emphasis was placed on selecting disease outbreaks outside the US to better understand the availability of data stream categories and ensure that identified useful data streams did not have a US centric bias. Other than Influenza, FMD and to a limited extent Cholera, most outbreaks identified described the disease spread within a specific country and did not often spread to other countries. Hence data streams evaluated were more representative of those used at local, state or national levels. We were able to find historical outbreaks to use as case studies for all diseases, however, in the case of West Nile virus, it was difficult to find information on an outbreak that would help us establish the timelines to include species jumps between birds and humans. Therefore, we simulated an outbreak for this disease and included the transition times between birds, mosquitos and humans.

Once the timelines for a disease were developed using the case studies, surveillance windows were defined and information for applicable data stream categories was collected for the duration of the outbreak. It was determined whether a particular data stream was available within the defined surveillance window, thus being deemed useful for either early warning or early detection of the disease.

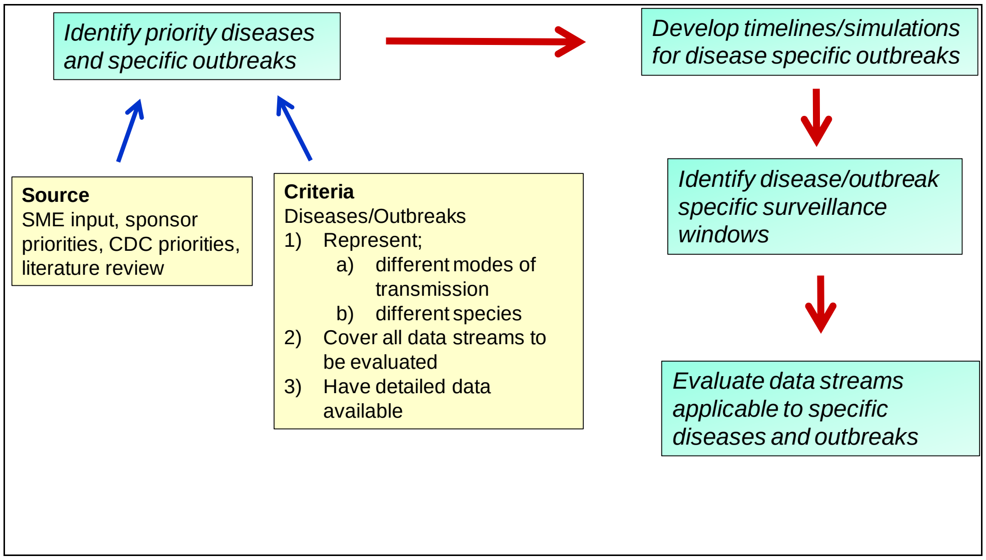


Figure 11: Overall approach to surveillance window based evaluation

The primary differences between this method of data stream evaluation and the MCDA based evaluation were that this method used a single metric for evaluation, time to indication, and a data stream category was evaluated by examining *specific* data streams that were available for each case study. A cross disease analysis was performed to identify data stream categories that consistently showed utility across syndromic classes of human diseases as well as across human, plant and animal diseases. A cross method analysis was performed between the surveillance window based evaluation and the MCDA-based evaluation to identify data stream categories that showed high utility for both methods.

Our first efforts were targeted at defining a clear approach to identifying surveillance windows, followed by a proof of principle demonstration using the 2009 Influenza H1N1 outbreak that originated in La Gloria, Mexico, followed by the remaining diseases. Over the course of evaluating various case studies, we refined the surveillance windows to facilitate a more robust evaluation of the data streams.

5.1 Defining Surveillance Windows

The concept of the surveillance window is tied to several factors, including the type of disease, operational needs and the specific characteristics demonstrated by the outbreak. We used a stepwise approach to defining disease specific surveillance windows;

1. Early dynamics of a large number of historical outbreaks were researched to generate timelines. Where appropriate, epidemiological simulations were used to supplement this data set.
2. The timeline was analyzed to determine the length of time between changes in "epidemiological state".
    a. A change in epidemiological state is marked by a change in the manifestation of the outbreak; either in terms of number of cases or in terms of geographical spread (see Figures 12 and 13). This may include the initial introduction of disease, a sudden increase in the number of cases, or the geographical spread to new areas.

   b. Changes are considered to correspond to specific operational needs and surveillance options, and the time point at which the specific change occurs is designated as the bound for a surveillance window.
3. In general, outbreaks go through several changes in epidemiological state. We initially considered each change, resulting in multiple surveillance windows for each disease. These windows also correlated with a surveillance goal. As our primary interest was early warning or early detection, the most useful data streams would need to fall in these windows to be of greatest use.
4. Data streams that were either used or could have been used due to their availability during the generated timeline were identified. If these data streams fell within each surveillance window, and provided both actionable and non-actionable information, they were deemed to have utility.

We tested our approach of defining surveillance windows and evaluating applicable data streams using the 2009 H1N1 influenza outbreak in Mexico. Figure 12 illustrates how we initially defined surveillance windows for this outbreak. Shown in this figure is the spread of H1N1 cases through La Gloria, Mexico, in February and March of 2009. There is an early period with a limited number of *"flickering" cases*, followed by a *rapid increase in case counts* and finally trailed by *peak cases*. All three of these "surveillance windows" are indicated in the figure. Each window has different surveillance signatures, surveillance best practices, and, very likely, different operational needs. The process used in defining surveillance windows from such information is explained in further detail below.

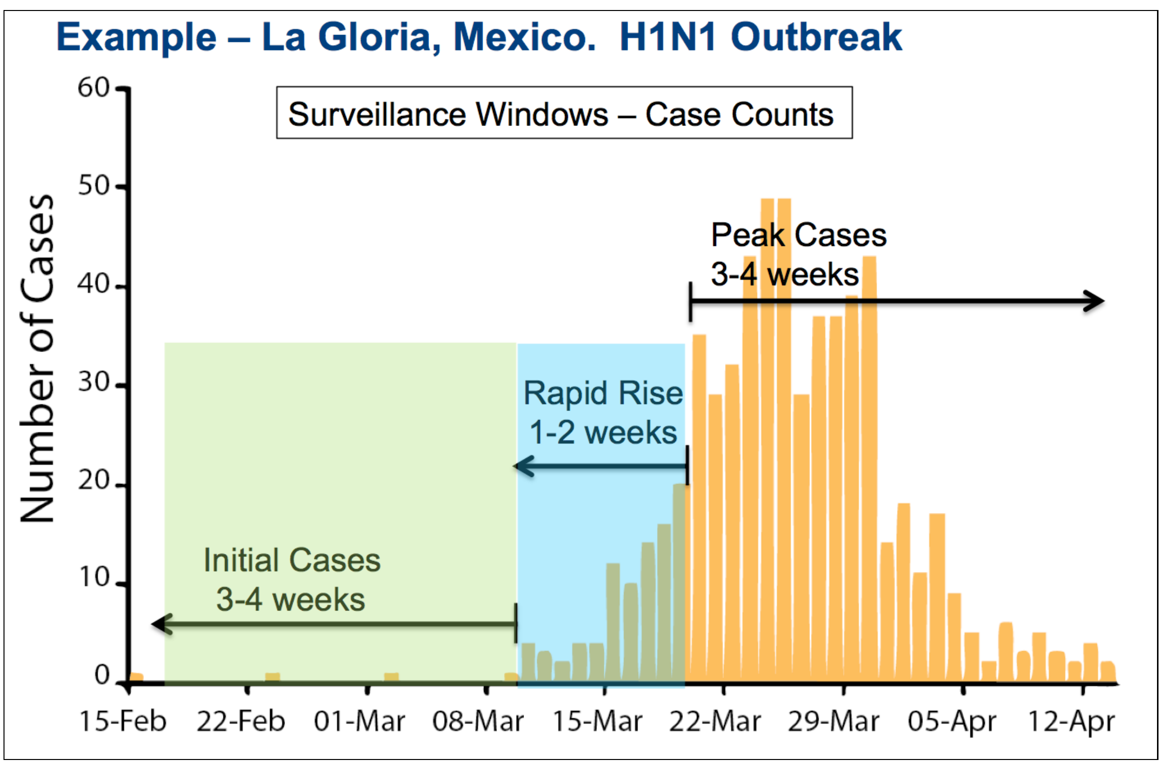


Figure 12: Surveillance windows for the H1N1 outbreak in La Gloria, Mexico are shown as described in the text. Case count graph adapted from Fraser *et al.*, *Science*, 324, 1557 (2009)

The index case is estimated as arising on February 15th, 2009, and small numbers of cases were recorded for the first three-and-a-half weeks. This period is our initial surveillance window, with the signature of this window being the “flickering” low number of cases. High sensitivity and broad population coverage are two of the necessary features of useful surveillance methods over this period. The “early warning” biosurveillance goal can be correlated with this surveillance window, though the use of this term may vary with the observer.

After this initial surveillance window, a rapid rise in the number of case counts occurred over a period of about one and a half weeks – a significant change in epidemiological state. This period is our second surveillance window, with the signature characteristic of this window being the rapid rise in the number of cases. While high sensitivity and broad population coverage are still useful requirements for surveillance systems in this surveillance window, a key feature is the ability to rapidly obtain actionable information. The “early detection” biosurveillance goal can be correlated with this surveillance window, though again the use of this term may vary with the observer.

The third surveillance window for the La Gloria outbreak is associated with all later/peak epidemic cases and occurred over a matter of 3-4 weeks. It begins at the case count inflection point (a change in epidemiological state) and continues until late in the outbreak. The key characteristic of this surveillance window is that some, typically incomplete, information is generally available regarding the outbreak – the need is for credible, actionable information. The “situational awareness” biosurveillance goal can be correlated with this surveillance window, though again, the use of this term may vary with the observer.

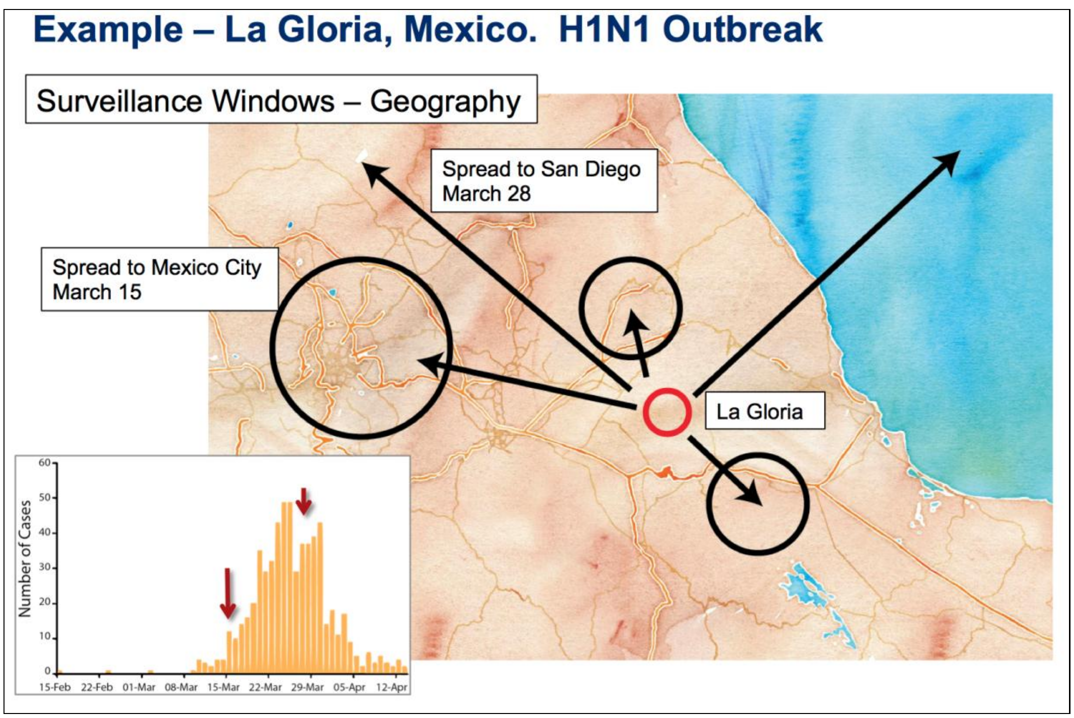


Figure 13: The geographic spread of disease can also result in a distinct and unique surveillance window. In this case, as described below, the spread to different regions in Mexico as well as the spread internationally was rapid and fell within or near previously identified surveillance windows. Case count graph adapted from Fraser *et al.*, *Science*, 324, 1557 (2009)

It is important to note that although we tried to directly relate specific surveillance windows with the surveillance goals of “early warning”, “early detection” and “situational awareness” the operational needs of any surveillance system must be considered in terms of geography and the goal of the end user. One user’s situational awareness may be another’s early warning. To address this issue, we also considered surveillance windows defined by the geographic spread of disease, as shown in Figure 13 – essentially treating the movement from area to area in the same manner as case counts were treated above. Conceptually, if a disease takes significant time to move from area to area, the value of late, but credible and actionable information (“situational awareness”) increases in value. In this case, we know that cases of H1N1 had spread to Mexico City very early in the outbreak timeline, essentially making the surveillance windows for regional spread roughly concurrent with those established for the town of La Gloria, potentially even a half week shorter for early detection. Similarly, international spread is evident very soon after the “rapid rise” surveillance window, suggesting a similar timeframe (1-2 weeks). Although a rapid rise in cases was observed for the country of Mexico as a whole near the end of April, the fact that the disease had spread worldwide before this time requires us to focus on the surveillance windows present in La Gloria, rather than developing windows specifically addressing geographic spread. It should be noted, however, that surveillance windows focused on geographic spread are a natural fit for several other outbreaks with different transmission characteristics – in particular with outbreaks that present strongly in animal hosts but where transmission to human hosts is of immediate concern.

The singular goal of the surveillance window was to provide a **visual** of the early stages of an outbreak during which certain data streams may be useful for the surveillance goals of early warning and early detection. While data streams that provide information early in the outbreak could be identified *without* the surveillance windows, it was possible to convey the disease and operations specific nature of the durations of time available and also perform comparative analyses across diseases using the visual of a surveillance window. Upon a deeper analysis of our approach to defining surveillance windows, there were certain limitations that became apparent in terms of a strong justification for the boundaries (both early and late) of the windows;

1) While the stages of the outbreak are being *correlated* with our goals, they are not meant to be an absolute time frame. However, the first depiction of surveillance windows appeared to convey that there were defined boundaries of time between the biosurveillance goals and the misinterpretation that early warning and early detection for a particular outbreak had very fixed start and stop times, when in reality, the boundary between those goals is not defined at all, but rather is very dependent on the progression of a particular disease through a population.
2) The placement of outer bound of the early detection window in correlation with the change in epidemiological state from a rapid rise in case counts or geographic spread to peak cases and spread, may be too late - it may appear to convey that we are indicating that “early detection” is possible even when a large population may have already become symptomatic and the rest are likely infected.

To address these limitations, we refined the boundaries of our surveillance windows in the following way;

- We defined a **single** surveillance window that encompasses the goals of early warning and early detection, and eliminated the artificial boundary between the two goals.
- We placed the left boundary of the window at the index case/index location minus the incubation period of the specific disease, to provide a better justification for when the window began.
- The right boundary was placed immediately before the rapid rise seen in the outbreak. This window captures two changes in epidemiological state –the first occurrence of the disease in a population and the change from "flickering cases" to a rapid rise. This outer bound captured the duration of time available *before* a majority of the population was already infected and therefore provided a more meaningful threshold after which mitigation may be too late. Within this surveillance window, the further left we are of an outbreak, the closer we are to the goal of early warning, and the further right we are, we transition to early detection.
- Both data streams available and in use prior to this rapid rise (pre-rapid rise window) and those available but not used were identified as being useful. The earlier time stamp they had, the closer they were to the early warning goal.
- Beyond the outer bound of the pre-rapid rise window, lie the goals of situational awareness and consequence management.

Figure 14 shows a comparison between the two iterations of surveillance windows for the 2009 H1N1 flu outbreak in La Gloria, Mexico. As can been seen the single surveillance window has a defined left and right boundary ansd is also shifted to the left on the timescale.

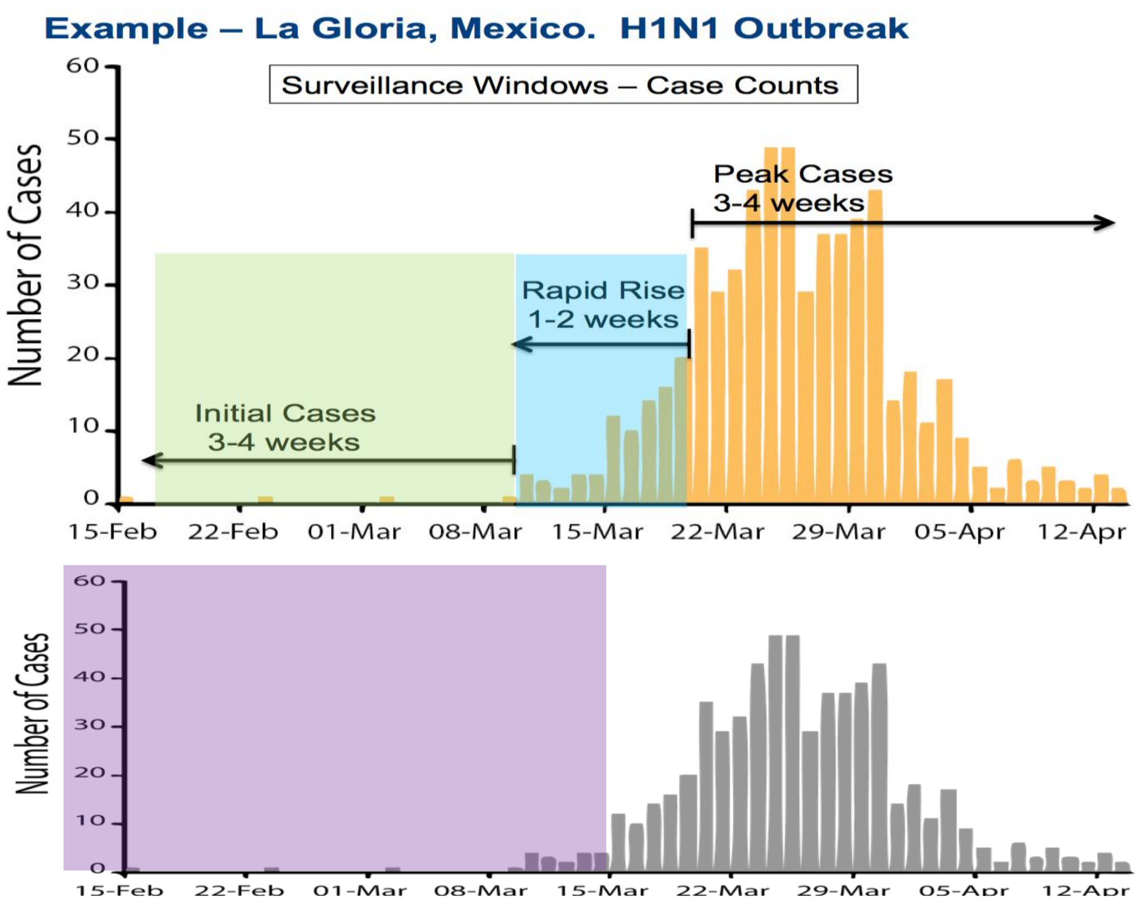


Figure 14: Modified surveillance windows for the H1N1 outbreak in La Gloria, Mexico

5.2 Evaluation of data stream categories using disease specific surveillance windows

Once the approach to defining surveillance windows was well defined, we proceeded with generating timelines, surveillance windows, collecting data stream information and evaluating data streams for disease specific case studies. Figures 15-23 depict 9 different case studies that spanned human, animal and plant diseases as well as represented bacterial and viral infectious diseases. Each figure depicts the timeline for a historical disease specific outbreak that started with an index case or index location and spread through the population over a period of time. The pre-rapid rise surveillance window was defined for each disease and is shown shaded in purple.

Below the timeline are identified time stamps for case study specific data streams (representative of the data stream categories described in earlier sections of this report) when they were first available shown in blue, transitioning to green when the data stream was recorded to be first used. Data stream categories, for which no information was found, are listed below the graph, under "Not Available". For every case study, the data streams that showed a time stamp that fell within the duration of the pre-rapid rise surveillance window were deemed to be useful. These included the data streams that were either used or were available during that time frame. The data stream categories are depicted in the figures in the order in which their time stamps first appeared on the timeline. Information collected to develop the case study timelines and data stream time stamps has been recorded and can be provided upon request. As mentioned before, this method of evaluation required the collection of information on very specific data streams. However, in order to facilitate cross disease and cross method analysis, we identified the specific data stream as its data stream *category*. For example, in case of the cholera case study in Haiti in 2010, Twitter was the specific data stream that information was collected about, but it was depicted as "Social Media" . A laboratory test for the cholera outbreak would necessarily be different than the one for a flu outbreak, but the specific data streams would both depict the category for "laboratory record". For each case study, examples of *specific* data streams found useful have been identified under each figure.

The primary benefits of the surveillance window approach for evaluating data streams is that it is based on extensive real data collected for each case study, and therefore offers high confidence in results of the evaluation. The method of evaluation is very simple and facilitates easy analysis of data streams used during outbreak timeline. The evaluation process also captures data streams that may be of benefit in the future if metrics surrounding the data stream are improved and identifies data streams of importance for containment of outbreak, post the pre-rapid rise surveillance window, but important for situational monitoring.

Some limitations of this method are that the results are based on "normal" outbreak progression and not deviations from the norm due to pathogen evolution, natural event overlap, etc. Thus a significant change in the progression may affect the pre-rapid rise window and therefore impact the utility of certain data streams. As the case studies are based on historical outbreaks, our results may be subject to change based on changes in availability of data streams over time.

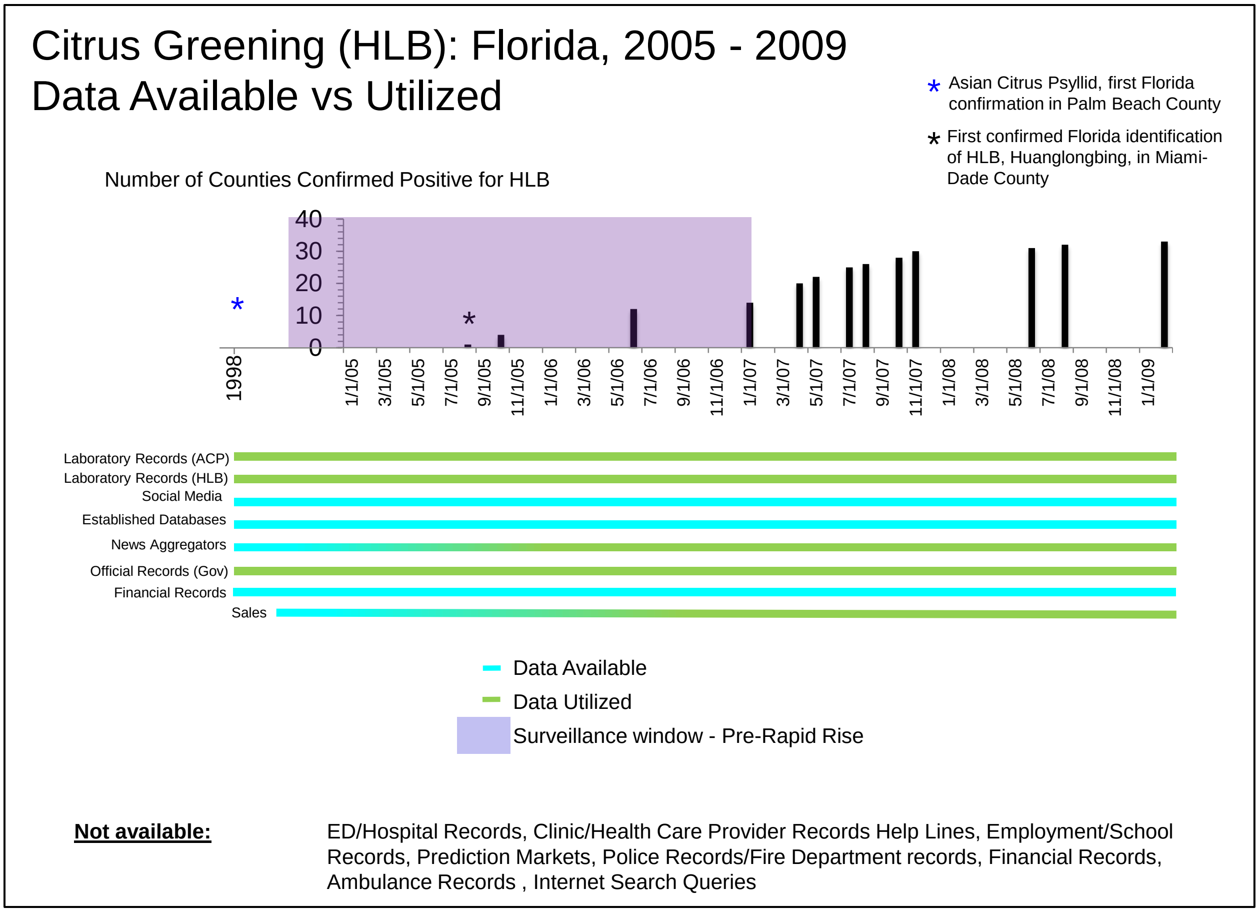


Figure 15: Data stream evaluation for Citrus Greening case study

Specific Data Streams Determined to Demonstrate Utility

Lab records and Official Reports: Florida Cooperative Agricultural Pest Survey (CAPS) teams (county specific) for the psyllid *D. citri* and for HLB. Surveys included either PCR or visual diagnosis. Reports through Division of Plant Industry (DPI) Florida Department of Agriculture and Consumer Services.

Social Media: Blogs (Citrus forum on Garden Web - website)

Established data bases: ex. Florida survey and mapping society - locations of plantations, climate databases

News Aggregators: ProMed

Financial records, sales: Retail trade in citrus (movement of psyllid) and plants (movement of HLB infection) orchards, nurseries, residential garden centers

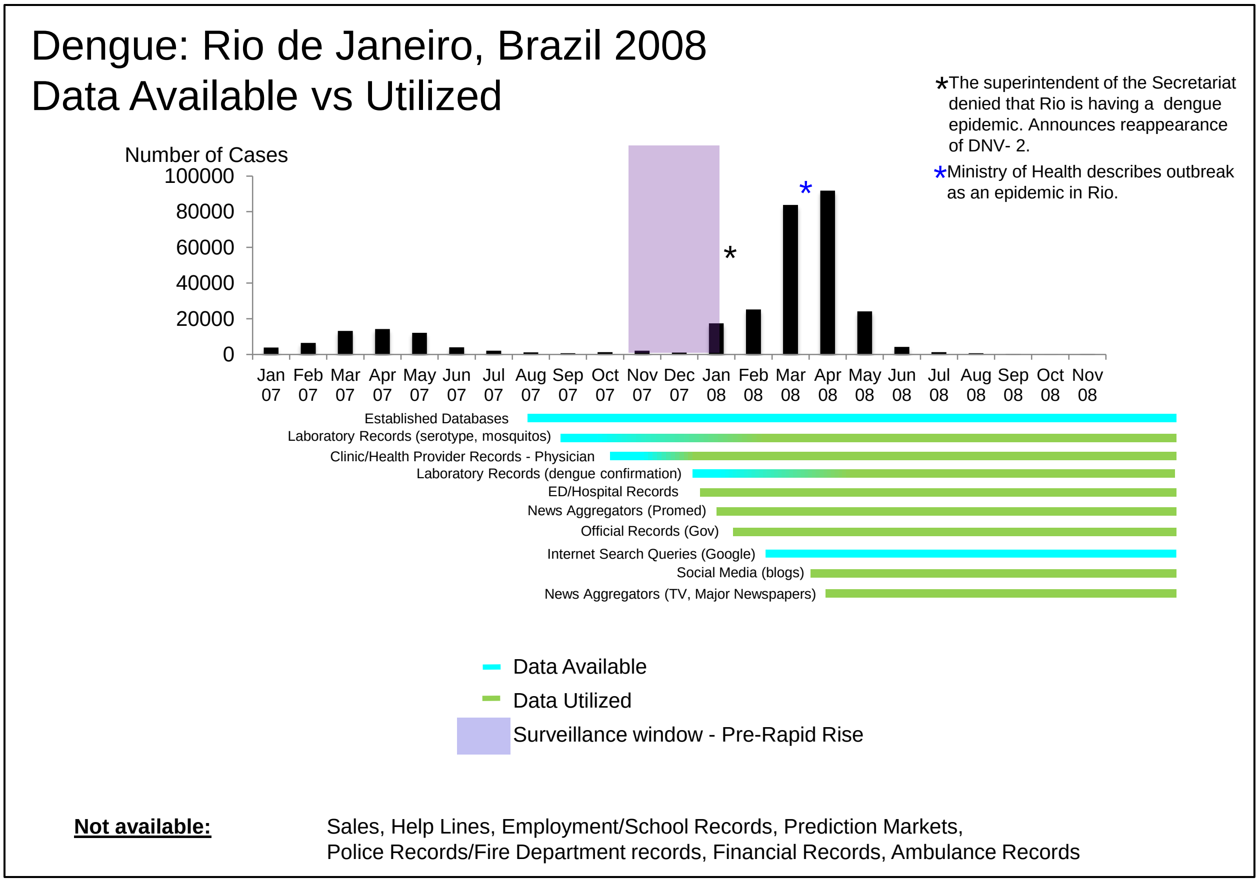


Figure 16: Data stream evaluation for Dengue case study

<u>Specific Data Streams Determined to Demonstrate Utility</u>

Established Databases: Instituto Brasileiro de Geografia e Estatística (IBGE), SIDRA Database)

Laboratory Records : Entomological surveillance: Ae aegypti eggs and adults**,** WHO Diagnosis of dengue in Brazil, Monitoring dengue virus serotype by PCR

Clinic/Health Care Provider Records: Clinical surveillance, SIH -SUS system for required reporting

ED/Hospital Records: The Hospital Information System of the Brazilian National Unified Health System (SIH-SUS)

News Aggregators : Promed - Jan 4 2008

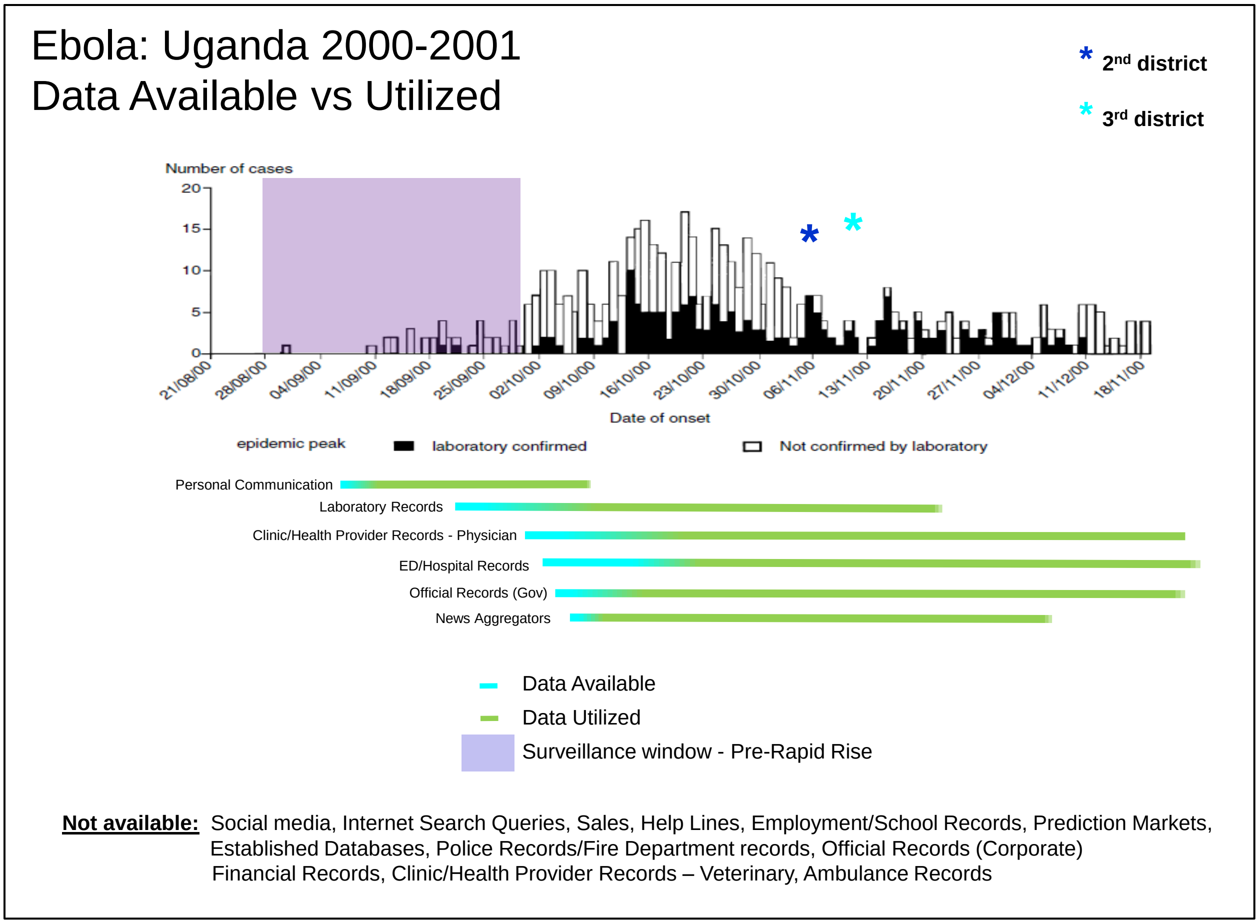


Figure 17a: Data stream evaluation for Ebola case study

<u>Specific Data Streams Determined to Demonstrate Utility</u>

Personal Communication: Surveillance teams (community and hospital based) for three districts

Laboratory Records: Temporary field laboratory at Gulu, National Institute of Virology (NIV), Johannesburg, South Africa

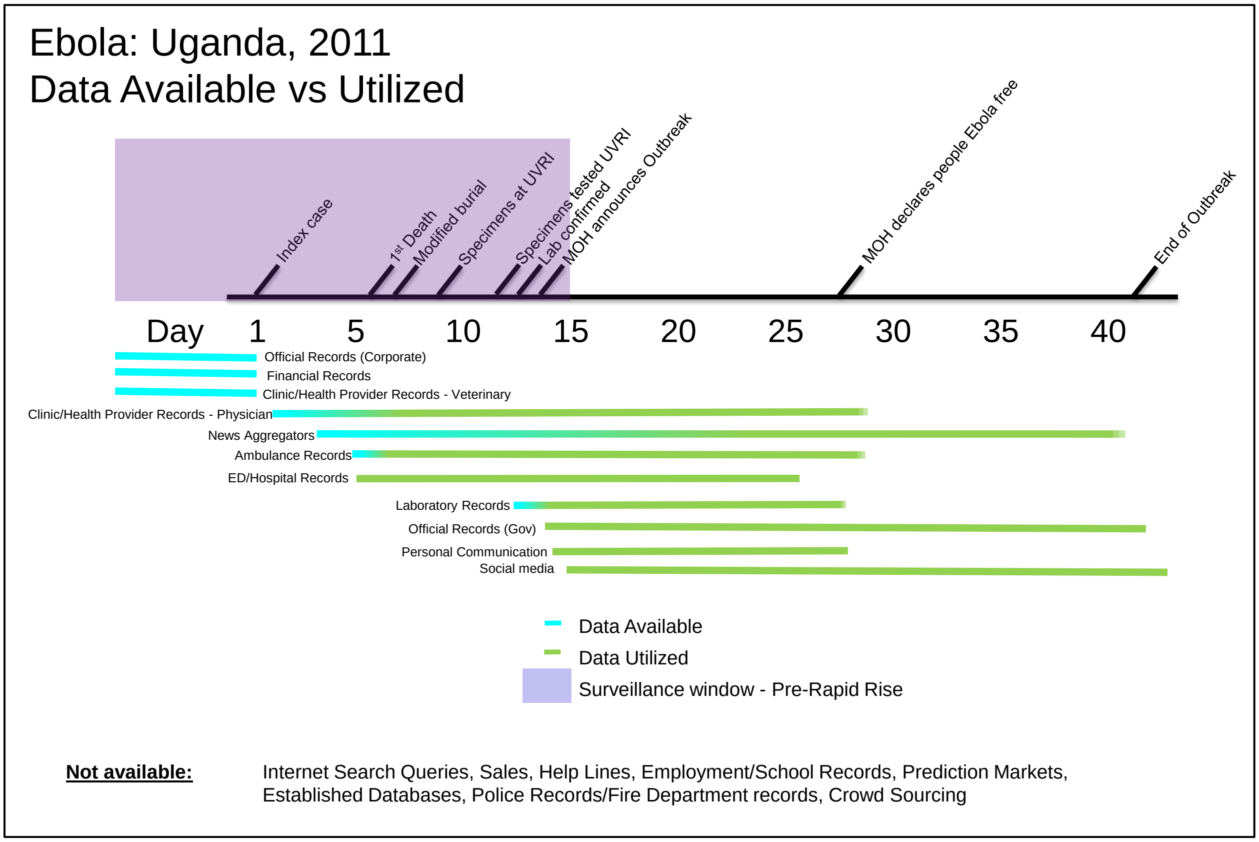


Figure 17b: Data stream evaluation for Ebola case study

Specific Data Streams Determined to Demonstrate Utility

Clinic/Health Provider Records : Kisakye Clinic in Zirobwe (Luwero district), Differential diagnosis performed, Bombo military hospital (Gombe Hospital)

News Aggregators: H5N1 Pandemic Information, Reuters,

Ambulance/EMT Records: Bombo military hospital

ED/Hospital Records: Bombo military hospital

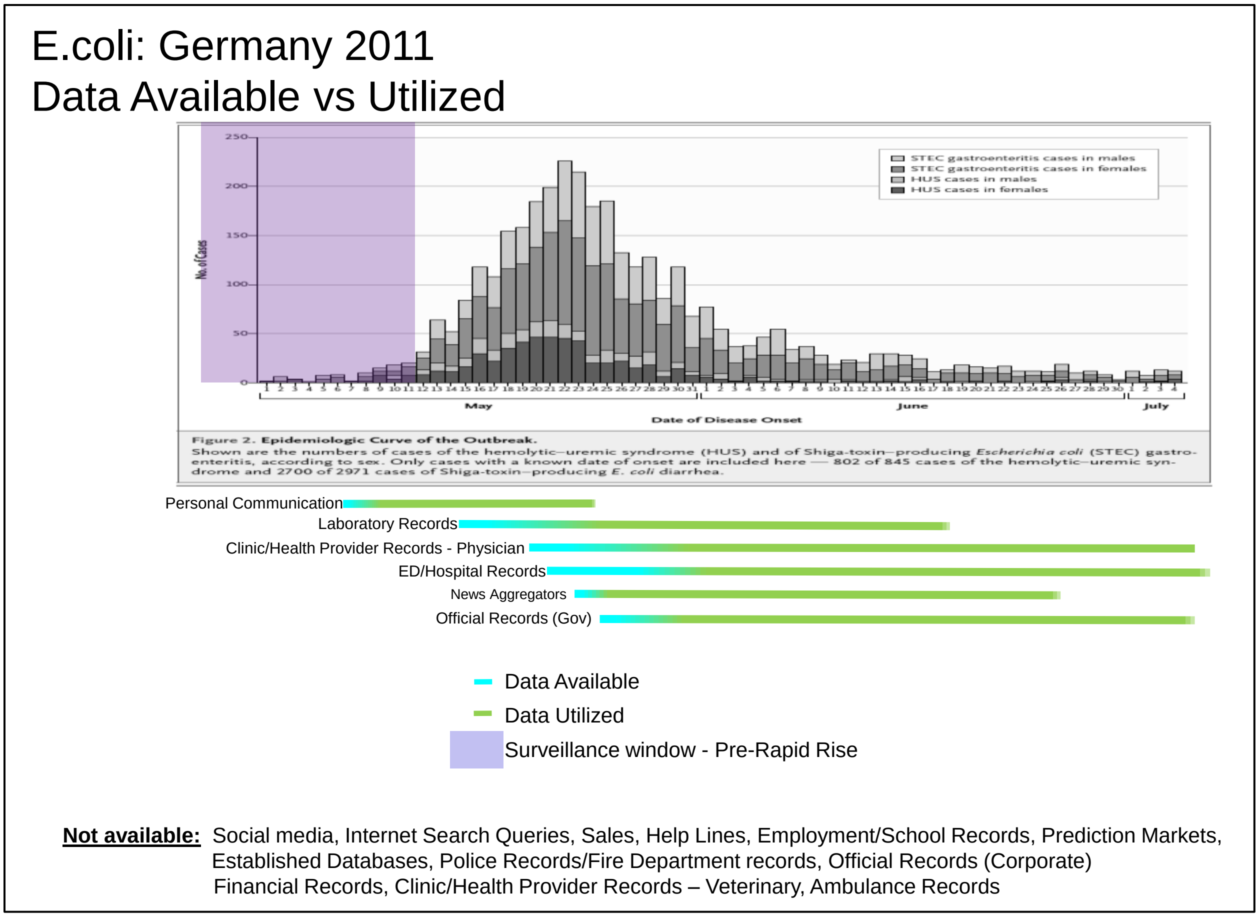


Figure 18: Data stream evaluation for *E.coli* case study

Specific Data Streams Determined to Demonstrate Utility

Personal communication: German health authorities, cases for diarrhea and hemolytic uremic syndrome. Number of cases of HUS or suspected HUS notified to local health departments and communicated by the federal states to the Robert Koch Institute (RKI). Sweden reported through the European Warning and Response System (EWRS)

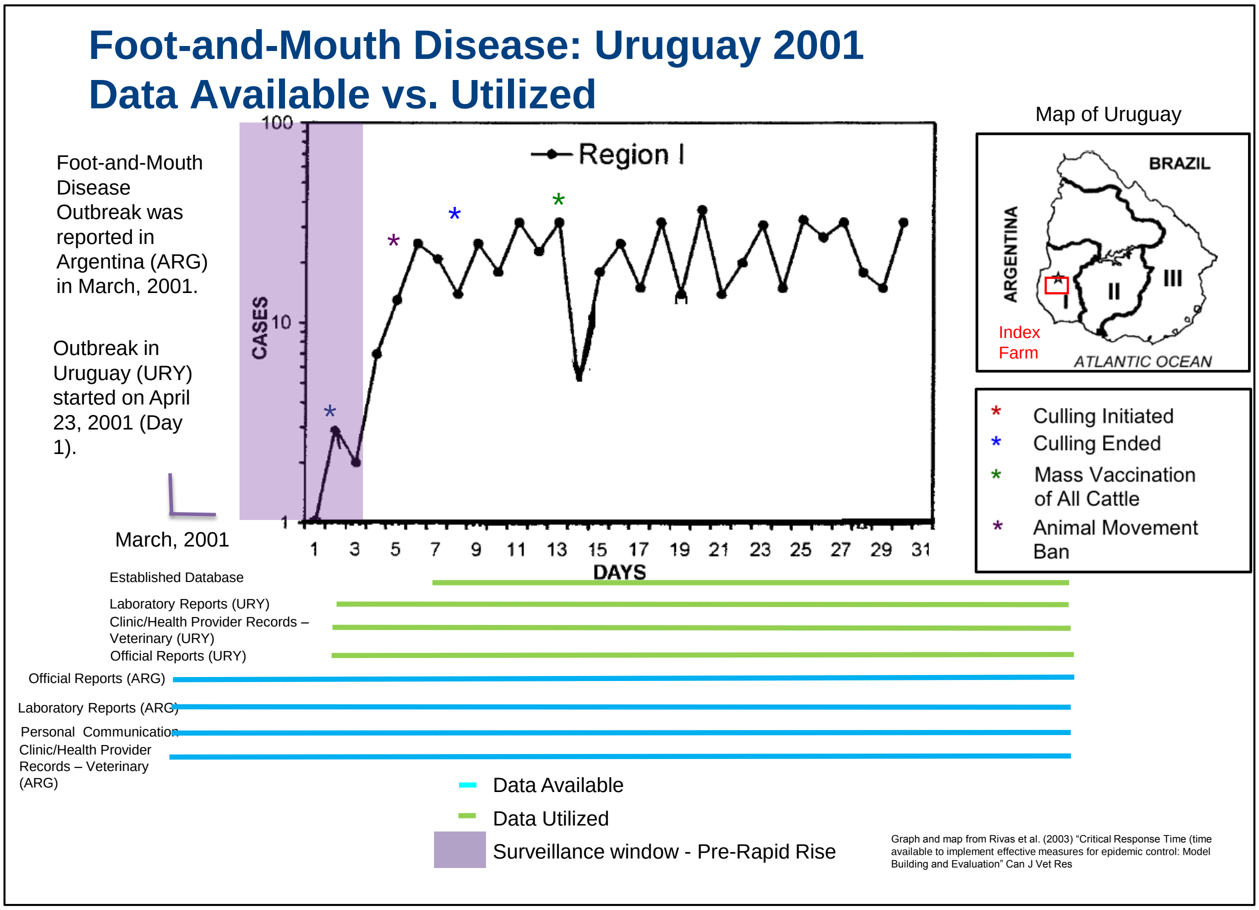


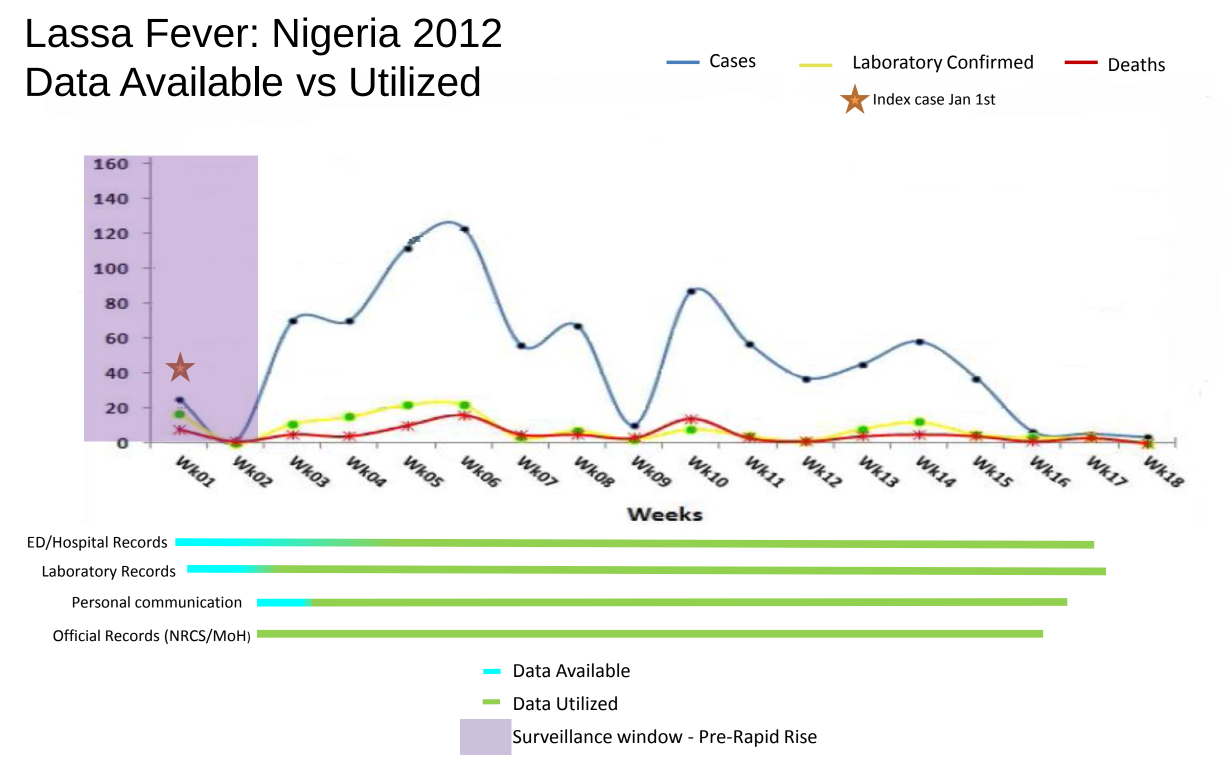
Figure 19: Data stream evaluation for FMD case study

Specific Data Streams Determined to Demonstrate Utility

Established Databases: FAO Emergency Prevention System (EMPRES)

Official Reports / Laboratory Records (Argentina): Generated by SENASA, a spanish acronym for The Agri-Food Health and Quality National Service. They issued reports and conducted the laboratory tests. SENASA website for Argentina

Official Reports and Laboratory Records for Uruguay: Generated by the Ministerio de Ganaderia Agricultura Y Pesca (MGAP), their ministry of Agriculture and Finishing



Not available: Internet Search Queries, Sales, Employment/School Records, Prediction Markets, Established Databases, Police Records/Fire Department Records, Financial Records, Clinic/Health Provider Records, Ambulance Records, Help Lines, Social Media, News Aggregators

Figure 20: Data stream evaluation for Lassa Fever case study

Specific Data Streams Determined to Demonstrate Utility

ED/Hospital Records: Ebonyi, Delta,

Laboratory Records: Irrua Specialist Hospital in Irrua, Edo State and the Central Medical Laboratory at the Lagos University Teaching Hospital (LUTH)

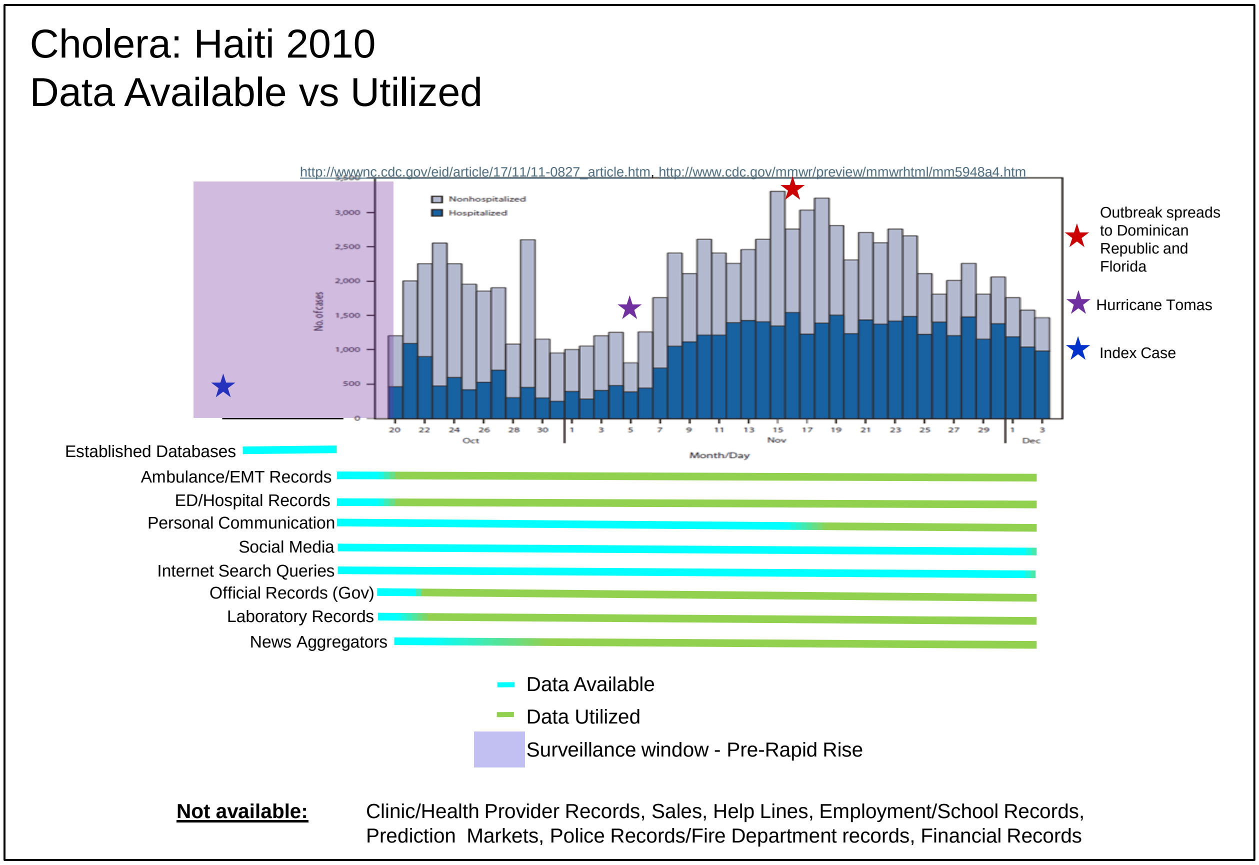


Figure 21: Data stream evaluation for Cholera case study

Specific Data Streams Determined to Demonstrate Utility

Ambulance/EMT records: Red Cross Ambulance services

ED/Hospital Records: Admission records from 2 hospitals in Artibonite

Personal Communication: Cell phone usage by families, CDC communication to Haiti Public Health lab

Social Media: Twitter

Internet Search Queries: Google Trends

Official Records (Gov): Haiti's Ministry of Public Health and Population (MSPP)

Laboratory Records: Haiti public health laboratory

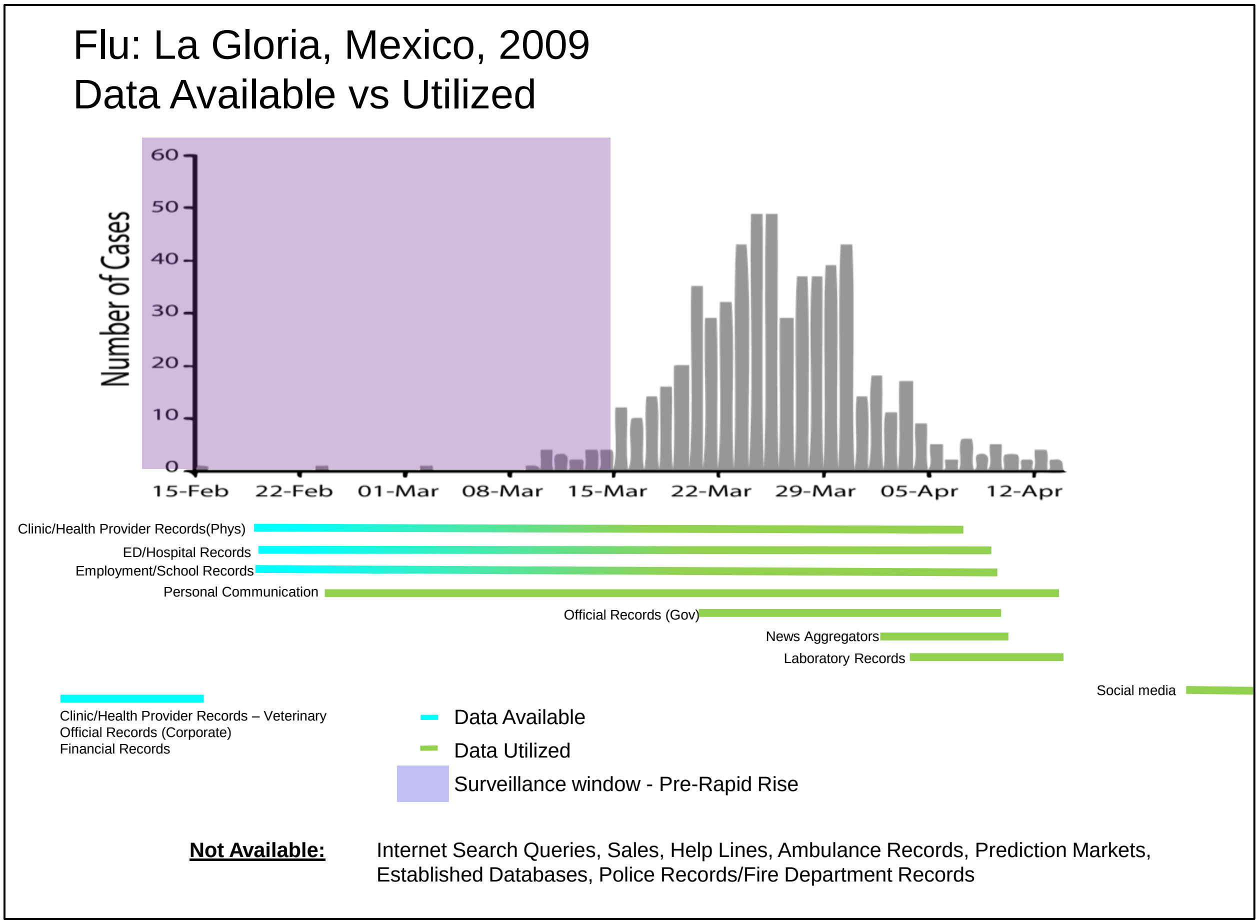


Figure 22: Data stream evaluation for Influenza case study

Specific Data Streams Determined to Demonstrate Utility

Clinic/Health Provider Records: La Gloria health clinic

ED/Hospital Records: hospitals in Perote and Jalapa, Center of Medical Specialties Rafael Lucio

Employment/School Records: schools never closed in La Gloria, but there were significant number of absences that may be in school records (New York Times Article). Teachers interviewed noticed absences

Personal Communication: Between families experiencing illness (as described in newspaper interviews)

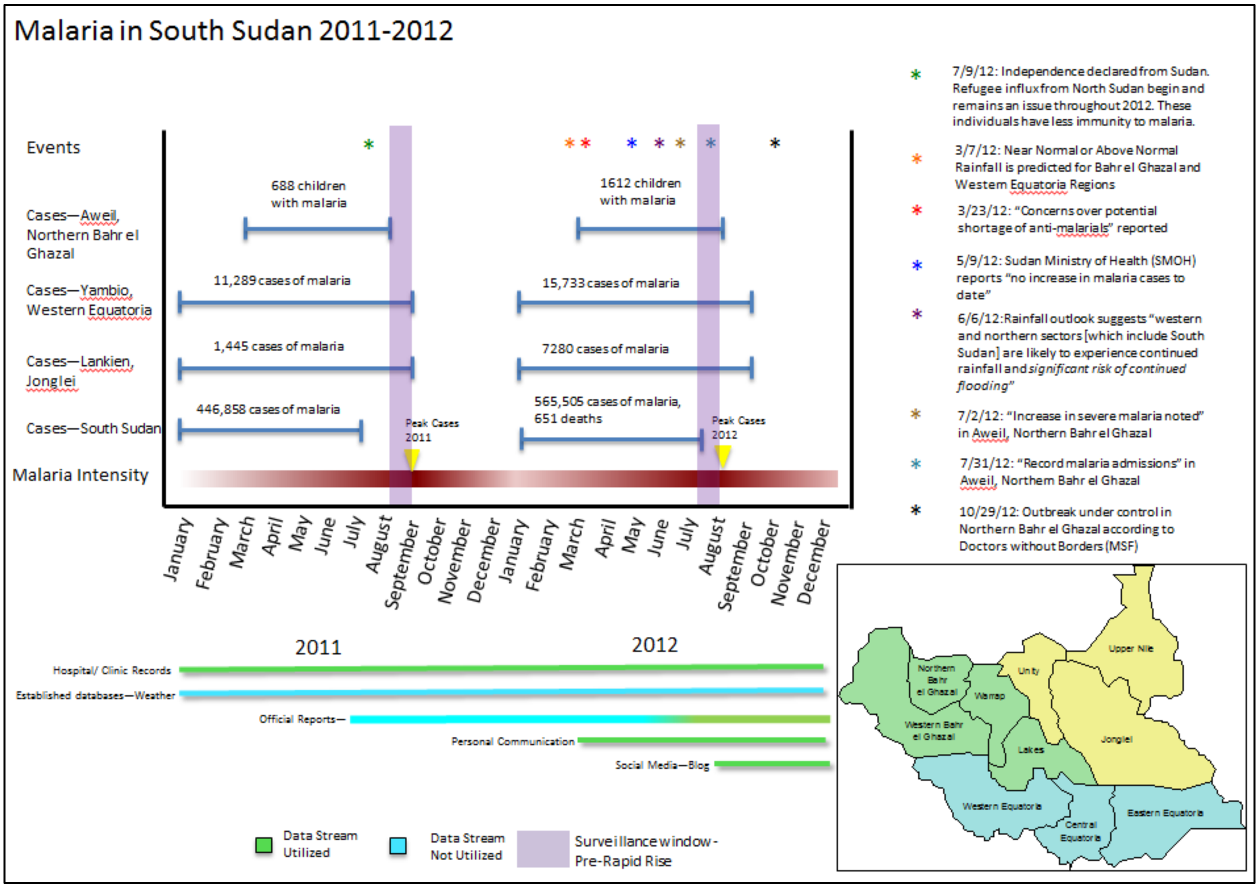


Figure 23: Data stream evaluation for Malaria case study

Specific Data Streams Determined to Demonstrate Utility

Established Databases (weather): Global Disaster Alert and Coordination System, Famine Early Warning Systems Network (FEWS NET) Africa Data Portal

Official Reports (weather): South Sudan Climis sponsored by the government of South Sudan) generates agro-meteorological reports, Famine Early Warning Systems Network (FEWS NET) Is sponsored by USAID and generates reports on weather predictions monthly.

Official Reports: United Nations Office for the coordination of Humanitarian Affairs—South Sudan ProMed, Outbreak warning provided by ProMed

Social Media: Bakhita Radio Blog, A blog run by a Catholic Radio Station that reports on happenings across South Sudan

Personal Communications: Meeting Minutes downloaded to the Google Cloud from public Health Organizations (ministry of Health, Doctors without Borders,etc.) in South Sudan

Clinic/Health Provider Records: Doctors without Borders maintains several clinics and used records to declare outbreak

Cross Disease Case Study Analysis of Results

Table 17 summarizes the surveillance windows defined for each of the case studies shown above in comparison with the average known incubation periods for each disease. Table 18 summarizes the results obtained from the evaluation of data stream categories using the 9 case studies. Comparisons across all case studies can be made for the data stream categories determined to demonstrate utility ("Used in Pre-Rapid Rise Window", "Available but not used in Pre-Rapid Rise Window"), and the data stream categories that were used for situational awareness (" Used in Post-Rapid Rise Window"). While this method does not provide a ranked list of data streams, it is easy to identify the data streams that appear earlier in the surveillance window and therefore can be correlated with the early warning goal of biosurveillance, and those that appear closer to the rapid rise of the outbreak and are therefore likely to be more useful for the early detection goal of biosurveillance. Using the single metric of time to indication for this method of evaluation, it was possible to group the data streams into three tiers with the earliest appearing ones and therefore the most useful data streams binned into Tier 1 (Table 19) Several interesting observations were made following a careful analysis of results presented in Tables 17 and 18;

1) While case studies spanned diseases like cholera that have an incubation period of about 2 days to Citrus Greening that have an incubation period 1-5 years, surveillance windows did not necessarily correlate with disease specific incubation periods. For example, influenza that has an incubation period of 3 days and Lassa fever, that has an incubation period of about 14 days, both appear to have long surveillance windows (28 and 33 days respectively). Further, dengue, which has an incubation period comparable to Influenza (5 days) has one of the longest surveillance windows (80 days). This suggests that the window of time available for early warning or detection of a specific disease is determined not only by the disease characteristics but also by operational characteristics, location, population and season, endemicity of disease, etc. Influences such as these could best be teased out by studying the same outbreak in different locations or different outbreaks in the same location.

   The Ebola case study provides a good example for the impact of two outbreaks, one in 2001 and one in 2011 in the same country, Uganda, in order to ensure that some of the more non-traditional data streams could be considered in our evaluation. Ebola outbreak analysis, when compared from 2001 to 2011, showed a significant shift in key data stream categories. Many of the categories listed in Table 18 as being available but not used until the post-rapid rise window in 2000 are shifted to the pre-rapid rise window in 2011, indicating the influence of a change in operations such as better hospital communication, containment procedures and containment practices (altered burial practices, key contacts communication/isolation) and increased awareness for the disease.
2) The duration of surveillance windows is varied and ranges from a few days to years. Despite this large variation, there were useful data stream categories that seemed common to most diseases; these were ones that relied on *specific, local and credible* information. The data stream

categories that showed consistent use and availability within the disease specific surveillance windows, regardless of human, animal or plant disease included **Laboratory Records, ED/Hospital Records, Clinic/Health Provider Records, Official Reports, and Personal Communication**. Interestingly, non-traditional data stream categories such as **News Aggregators and Social Media**, while not used for decision making, did show availability and therefore utility within the surveillance windows.

Case Study	Window Duration	Incubation Period
Cholera - Haiti (2010)	10 days	2 days
Citrus Greening - Florida, USA (2005)	1-5 years	1-5 years
Influenza - Mexico (2009)	33 days	3 days
Ebola - Uganda (2001)	37 days	7 days
Ebola - Uganda (2011)	* single case outbreak, ~ 23 days	7 days
Lassa Fever -Nigeria (2012)	28 days	14 days
FMD - Uruguay (2001)	6 days	3 days
Dengue - Brazil (2008)	80 days	5 days
Foodborne E.coli - Germany (2011)	16 days	4 days

Table 17: Surveillance window durations and incubation periods for each disease

Case Study	Used in Pre-Rapid Rise Window			Available but not used in Pre-Rapid Rise Window	Used in Post-Rapid Rise Window
	1	2	3		
Cholera - Haiti (2010)	Ambulance/EMT Records, ED/Hospital Records	Laboratory Records		Official Records (Gov), News Aggregators, Personal Communication, Internet Search Queries, Social Media, Established Databases	Official Records (Gov), News Aggregators, Personal Communication, Internet Search Queries, Social Media
Citrus Greening - Florida, USA (2005)	Laboratory Records, Official Records (Gov)	News Aggregators * all 3 were early warning indicators	Sales	Social Media, Established Databases, Financial Records	
Influenza - Mexico (2009)	Clinic/Health Provider Records (Physician), ED/Hospital Records, Employment/School Records	Personal Communication		Official Records (Corporate), Financial Records, Clinic/Health Provider Records (Veterinary)	Official Records (Gov), News Aggregators, Laboratory Records, Social Media
Ebola - Uganda (2001)	Personal Communication	Laboratory Records			Clinic/Health Provider Records (Physician), ED/Hospital Records, Official Records (Gov), News Aggregators
Ebola - Uganda (2011)	Clinic/Health Provider Records (Physician), News Aggregators, Ambulance/EMT Records, ED/Hospital Records	Laboratory Records	Personal Communication, Official Records (Gov)	Official Records (Corporate), Financial Records, Clinic/Health Provider Records (Veterinary)	Social Media
Lassa Fever -Nigeria (2012)	ED/Hospital Records, Laboratory Records		Official Records (Gov)		Personal Communication
FMD - Uruguay (2001)	Clinic/Health Provider Records (Veterinary-URY), Official Reports (URY), Laboratory Records (URY)			Personal Communication, Official Records, Laboratory Records (ARG), Clinic/Health Provider Records (Veterinary-ARG)	
Dengue - Brazil (2008)	Clinic/Health Provider Records (Physician), Laboratory Records	ED/Hospital Records, New Aggregators (Promed)		Established Database	Personal Communication, Official Records (Gov), Social Media, News Aggregators (TV, Newspapers)
E.Coli - Germany (2011)	Personal Communication				Laboratory Records, Clinic/Health Provider Records (Physician), Official Records (Gov), News Aggregators, ED/Hospital Records

Table 18: Cross disease case study analysis of results of data stream evaluation

Cross Method Analysis of Results

To identify the data stream categories that demonstrated utility using both the MCDA based and surveillance window based approaches, the list of ranked data stream categories obtained with the MCDA approach were compared to the tiers of data streams obtained through the surveillance window approach. As these tiers of data streams were obtained using the single metric of time to indication and therefore related to the goals of early warning and early detection, we compared these tiers to the ranked list of data streams obtained for the same goals with the MCDA approach. The MCDA ranked list was also grouped into three tiers, with the highest ranked data streams in Tier 1. Results are shown in Table 19. Overall, the distribution of data streams among the three tiers was fairly similar between the two methods, which provides higher confidence in the data streams that were determined to have the highest utility; **Laboratory Records, ED/Hospital Records and Clinic/Health Provider Records** showed the highest utility using both methods, while Prediction markets, Police/Fire Department records and Financial records showed the lowest utility. There were minor differences in Tier groupings for all the other data stream categories. Non-traditional data stream categories such as **Social Media and News Aggregators** remained in the middle tier suggesting moderate utility. The most disparity in rankings was with **Internet Search Queries**, which was ranked very high with the MCDA approach, but was binned to the least useful tier with the Surveillance Window approach.

There is likely a bias in the overall rankings for these data stream categories due to the availability of more information for many of the top ranked categories, as well as demonstrated use for a significant period of time. Therefore there is higher confidence in values that were assigned to them. However, the middle tier of data stream categories will be important to consider along with the top tier for this reason.

It is also important to note that such a simplified analysis does not reveal disease specific and operations specific data streams that may be useful, and that can be identified only through the analysis of the individual case studies. For example, **Sales** are an important data stream in the context of plant disease surveillance due to the accessibility of information and lack of privacy considerations as compared to human disease surveillance. Another example is the early use of **Employment/School Absenteeism Records** for Influenza surveillance. We also identified disease specific data streams currently not in use that could be exploited for faster outbreak detection. For example, in the FMD case study in Uruguay in 2001, our analysis indicated that Official reports, Laboratory reports, Personal communication and Clinic/Health Records (Veterinary) from the neighboring country of Argentina were available and could have been used to aid in biosurveillance efforts in Uruguay. During the Influenza disease progression of 2009 in Mexico, Clinic/Health Provider Records – Veterinary, Official records (corporate) and Financial Records were all available early or before the initial case but were not utilized in a biosurveillance context. Established databases, such as Weather pattern data, Toxnet, etc., appeared consistently across diseases in a both pre- and post-rapid rise window timeframes as available but not utilized effectively for biosurveillance efforts. Finally, we identified data stream categories that were consistently used for post-rapid rise surveillance correlated with situational awareness/monitoring and mitigation. These included News Aggregators, Official Records (government) and Social media.

Such data streams, while not showing high utility in overall analysis should not be discounted for disease specific surveillance. The data stream categories deemed most useful through both methods should be included in a disease surveillance system, and additionally, those deemed useful for specific diseases should be included to support better surveillance for that disease. For example, Laboratory Records, ED/Hospital Records and Clinic/Health Provider Records can be used regardless of the disease being monitored, but School/Employment Records could be used for Influenza, and Sales could be used for plant disease surveillance. Such a strategy would allow for a more efficient and streamlined surveillance system OR to decrease the amount of data that an analyst would have to focus on.

	Surveillance Window Ranking	MCDA Ranking
Tier 1	Laboratory Records	Internet Search Queries
	ED/Hospital Records	ED/Hospital Records
	Clinic/Health Provider	Clinic/Health Provider
	Official Records (Gov)	Laboratory Records
	Personal Communication	News Aggregators
Tier 2	Ambulance/EMT Records	Help Lines
	News Aggregators	Social Media
	Social Media	Ambulance/EMT Records
	Sales	Personal Communication
	Established Databases	Official Records (Gov)
Tier 3	Internet Search Queries	Sales
	Financial Records	Police/Fire Department Records
	Employment/School Records	Employment/School Records
	Help Lines	Financial Records
	Police/Fire Department Records	Established Databases
	Prediction Markets	Prediction Markets

Table 19: Cross method analysis of data streams evaluation results

5.3 Evaluation of data stream categories for zoonotic diseases

While information on early disease dynamics for zoonotic diseases can be obtained for one or the other species from historical outbreaks, producing a timeline that includes the occurrence of species jump is very difficult (e.g West Nile virus). As a result, the ability to identify windows of time prior to human infections (early warning/prediction) where animal and vector surveillance could be evaluated is limited. This is where simulating a zoonotic disease outbreak becomes useful. We used the case study of a West Nile virus outbreak that jumped from birds to humans and simulated a timeline to include the duration between the first bird case and the first human case. Figure 24 shows the simulation that follows a potential outbreak scenario in a small suburban town of roughly 2,500 people spread over 1 square kilometer. It was assumed that there were roughly 400 birds moving around in the spatial domain as well as 90,000 mosquitos. The relative prevalence and bite preferences of three different mosquito species were modeled based on [Molaei 2006] and our own work along with 25 difference species of bird (each with different disease parameters and mortality rates based on [Komar 2003a]). Bird and mosquito species relative abundances were representative of New England. The basic structure of the bird and mosquito models was based on [Wonham 2004], though our simulation used an agent-based representation of birds and humans and a multi-community ODE representation of mosquitos (validated against agent-based mosquito runs of smaller size). Mosquito density was assumed to be about twice normal mosquito densities providing a richer environment for an outbreak to occur. Statistics were based on data collected over 50 runs.

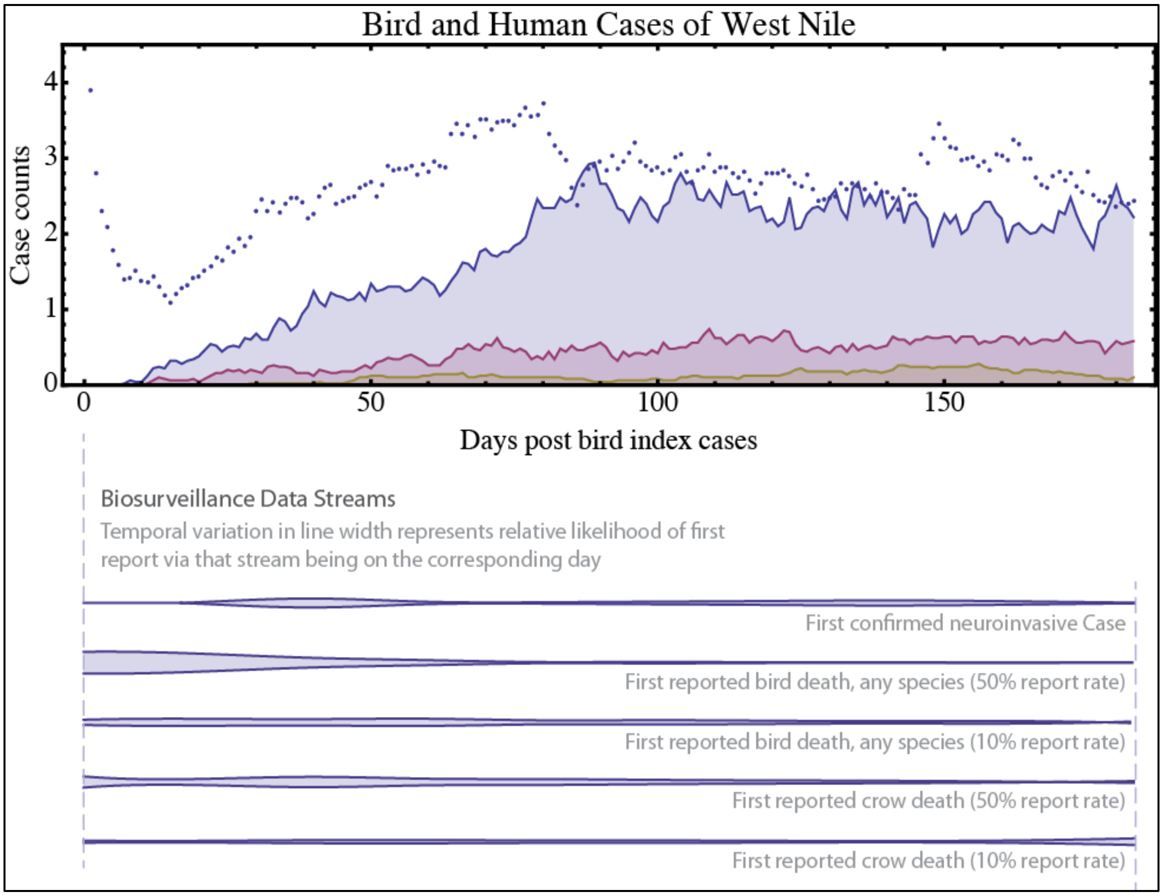


Figure 24 Top: Evolution of a West Nile outbreak beginning with four infected birds. Each data point represents an average of over 50 runs. Each run simulated West Nile transmission for six months through a population of roughly 400 birds, 2,500 people, and 90,000 mosquitos. Blue dots represent the number of infected birds. The solid curves indicate: (blue) the number of asymptomatic human cases; (red) the number of febrile cases; and (yellow) the number of neuroinvasive cases. Note that outbreaks of West Nile Fever often don't behave in the same way that outbreaks of, for example, influenza do. Rather than an exponential increase in case counts, outbreaks tend to smolder at low levels for long periods, slowly racking up cases. **Bottom:** The relative likelihood of the first potential report of a West Nile outbreak being available via different data streams is shown as a function of the elapsed duration of the epidemic. For example, there is a reasonable likelihood that the first neuroinvasive human cases would be observed within the first 30-60 days of an outbreak. (It is assumed that 100% of neuroinvasive cases will be properly attributed to West Nile.) Testing times are excluded here (i.e., if a test typically takes two weeks to return a confirmation, the curve should be shifted right 14 days). By contrast, one might expect to see indications of an outbreak in the first few weeks if bird deaths are being monitored, but only if all species are monitored and one can expect there to be a 50% chance that a dead bird is found and reported. If only crow deaths are monitored and the reporting rate drops to 10% (last curve), the data stream is much less effective in the early identification of an outbreak.

Evaluating data streams

The West Nile virus today has a nearly global distribution. As mosquitos, the primary West Nile vector, have very low mobility (on the order of 50m per day) [Trpis 1986], the long-range spread of the virus is accomplished primarily by the movement of reservoir species (*e.g.*, migratory birds). As restricting the movement of these species is infeasible, prevention and control efforts have necessarily focused on local control of the vector species. Two non-exclusive approaches are frequently used: source reduction (through improved sanitation and the management of standing water) and chemical control (using insecticides that target either larval or adult mosquitos). The approaches can work in tandem and both can be used as either preventative or responsive measures.

Data streams that indicate the presence of West Nile virus in vector, reservoir, or incidental host populations and/or the virus's spatial distribution are useful in helping public health officials determine where and when mitigative resources should be mustered to prevent or control an outbreak. The paragraphs below describe some of the data streams that can contribute to early warning or detection of West Nile virus outbreaks in humans, as they are focused on animal and vector surveillance. It is not prudent to place a surveillance window on a simulation, because unlike timelines obtained from actual historical outbreaks, a simulated timeline does not take into account real world operational influences on the progression of the disease. However, it can be inferred from the timeline that **both animal and vector surveillance** would be useful due to the fact that birds do get infected before humans. A number of different surveillance methods are used targeting various host and vector species in different ways. The bulk of the information in this section comes from [CDC 2003].

1) Avian surveillance

- Dead bird surveillance – Community involvement in reporting the presence and, preferably, the geolocation of dead birds is one of the sensitive surveillance tactics with respect to early detection of the virus in most areas. Corvids are an especially good indicator because of their high susceptibility nearly 100% mortality rate.
- Live bird surveillance – The use of either captive sentinels or the surveillance of wild bird populations can be indicative of the West Nile virus.
- In both cases, conclusive attribution to the West Nile virus requires serological testing of specimens which can take up to three weeks and be prohibitively expensive.

2) Mosquito surveillance

- Dip or trap sampling of mosquito larvae or adults
- Vector surveillance can be the most accurate tool for quantifying outbreak spread and severity. It can also provide an early indicator of the introduction of West Nile virus into an area.
- Labor intensive and expensive.

3) Equine surveillance

Monitoring for the presence or spread of West Nile virus among equine populations has both pros and cons.

- Horses are a potentially great sentinel species because of their high exposure to mosquitos and frequent observation by owners.
- They are also often widely distributed throughout a region allowing for both increased sensitivity and spatial resolution.
- Widespread vaccination can significantly decrease incidence and surveillance sensitivity.
- Testing is expensive, time-consuming, and the cost is usually borne by the owner.

Human clinical surveillance is limited primarily due to the limited number of neurological cases that present and the non-specific nature of symptoms associated with this disease that result in fewer people visiting hospitals or clinics.

- Laboratory tests (ELISA) can be used to conclusively attribute cases of neurologic disease in both humans and horses to the West Nile virus.
- The same is true for febrile clinical cases, though the cost is typically prohibitive given the high background level of influenza-like illness that would trigger testing.

6.0 A Review of Data Integration Algorithms

The threats posed by natural epidemics and bioterrorism have necessitated the development of disease detection systems. At the core of these systems are algorithms that can parse through and analyze health related data for early warning or detection of an outbreak, thus generating alerts for public health officials to investigate. Currently, most operational systems tend to analyze multiple univariate data feeds simultaneously, that is, there are multiple data feeds each being analyzed with a single algorithm. Integration of data streams can potentially increase the capability of these systems by increasing sensitivity of the analysis.

Homeland Security Presidential Directive 21 defines biosurveillance as meaning "the process of active data-gathering with appropriate analysis and interpretation of biosphere data that might relate to disease activity and threats to human or animal health – whether infectious, toxic, metabolic, or otherwise, and regardless of intentional or natural origin – in order to achieve early warning of health threats, early detection of health events, and overall situational awareness of disease activity" (HSPD-21). The majority of currently operational systems are not able to incorporate diverse types of data and, instead, rely solely on syndromic health data gained from hospital or clinic records, over the counter (OTC) sales, or absenteeism records. Thus, most biosurveillance conducted in the United States is primarily syndromic surveillance. A study conducted by Buehler et al. in 2008 surveyed state, territorial and selected local health departments on their use of syndromic surveillance. They found that 84% of departments used emergency department visits, 44% used OTC sales, 37% used call data from poison control centers, and 35% used school absenteeism records. Given the focus of public health departments on syndromic surveillance, most of the algorithms being used to facilitate early warning or detection of an outbreak are designed to use syndromic data and do not incorporate more diverse types of information into their analyses.

It is important to understand how algorithms use data to generate reports. While algorithms are data agnostic, in the context of biosurveillance they have been designed to use syndromic data. A single data stream may contain multiple data feeds. For example, a hospital emergency department data stream may contain multiple feeds each detailing the time series of a specific syndrome (Figure 25).

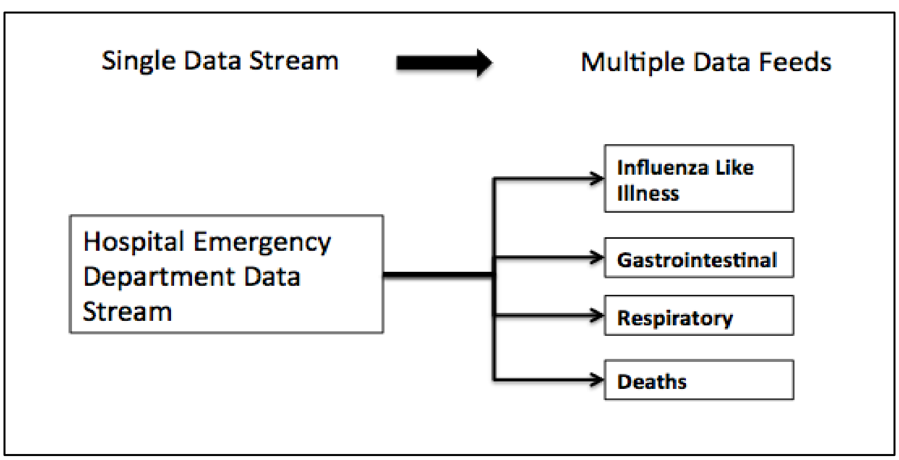


Figure 25: Data feeds coming from a single hospital

As shown in Figure 26, before the algorithm can use the syndromic data, it needs to be pre-processed so that periodic effects (day of the week effects, seasonal trends, etc.) in the data can be removed. Pre-processing is also required to homogenize the structure of different data feeds that are inputted into the algorithm. Generally, it is easier to pre-process data feeds from the same data stream, than it is to pre-process data feeds across many different data streams. Additionally, the statistical method being used to analyze the data (algorithm) needs to be parameterized with the appropriate historical data in order to properly observe the trends being looked for. For example, in the case of a flu outbreak detection algorithm, this can mean providing it with historic, regional data for flu over the past three years.

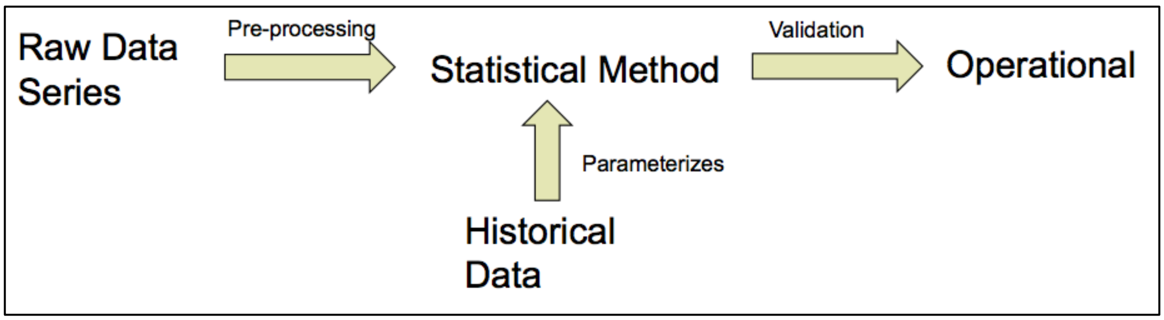


Figure 26: Raw data processing flow

The concept of integration often gets used ambiguously when discussing biosurveillance. It can be understood in three ways. The first is often used in the sense of aggregation of data from multiple sources into a larger data set. Difficulties in data aggregation arise from the need to pre-process each data feed so that they can be appropriately structured and combined. As can be imagined, different data sets from different hospitals, for example, need individualized pre-processing techniques to be applied since it is possible that the data sets are not in the same format or structure. Examples of this type of integration include aggregation of emergency department records from different hospitals within the same county, or OTC sales from different stores located in a single region.

The second common meaning of integration in the context of biosurveillance is the timely sharing and exchange of information amongst different levels of government. This includes reciprocal sharing across local, tribal, state, and federal as well as amongst federal agencies. Many barriers to this type of integration involve policy, organizational structures, and inadequate infrastructure.

The third meaning of integration is the ability to conduct analyses using a synthesis of multiple, different data stream and types (multivariate statistical analysis). This type of integration relies on the assumption that there are relationships amongst certain data types, that when combined can result in a more sensitive analysis. The relationship between Internet search queries, OTC sales, and emergency department records are an example of this. When monitoring a population for an outbreak, spikes in OTC sales and Internet search queries are likely to precede a spike in emergency department cases. Thus, if the appropriate multivariate statistical methods are used, it may be possible to trigger an outbreak alarm quicker than if each data stream is analyzed separately. Barriers to implementing this type of integration result from the lack of a "gold" standard data set in which to parameterize the algorithms. Additionally, stemming from this lack of a “gold” standard data set, the statistical methods used by the algorithms to analyze multiple data streams need to be validated.

We surveyed the use of algorithms to integrate multivariate signals for early event detection. Other types of biosurveillance algorithms, while important and relevant to early event detection and biosurveillance, were beyond the scope of this project. These include algorithms that are involved in data aggregation, algorithms for pre-processing (Exponential Smoothing, ARIMA/ARMA, wavelet methods, single spectral analysis), and algorithms for forecasting. The models are important for pre-processing the data and establishing the expected baseline with which actual disease incidence is compared to. Additionally, there are multivariate early event anomaly detection algorithms that are variations on the same statistical method. While not discussing every single variation we describe the statistical methods involved and how they are implemented in the algorithms of currently operational biosurveillance systems.

Monitoring for early event detection of an outbreak can take two forms. The first is monitoring multiple univariate data feeds for an anomaly. An algorithm is analyzing a single data feed. There are several advantages to this approach, it is easy to implement and interpret but it can be less sensitive. The second is type of monitoring is when a single algorithm monitors multiple data streams. This approach allows for greater sensitivities by taking relationships amongst data feeds into consideration but it is much more difficult to implement and interpret and can lead to excessive false alarms being raised. Currently, the majority of biosurveillance systems analyze data using multiple univariate streams. There are no multivariate analyses currently being used in these systems. A comparison of the two forms of data integration are shown in Figure 27

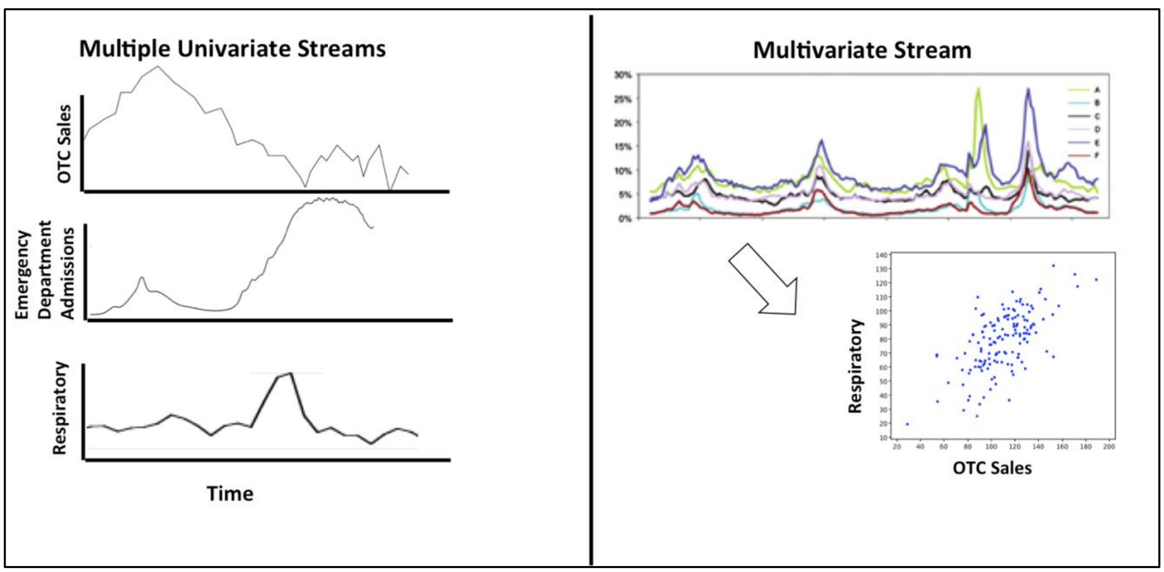


Figure 27: Forms of data integration for anomaly detection. Image From:http://www.eht-journal.net/index.php/ehtj/article/viewFile/11907/17325/44991

Most biosurveillance early event detection algorithms are adaptions of statistical methods that were developed for statistical process control (SPC). SPC was developed as a quality control monitoring method using statistical techniques primarily for industrial applications. Most of the techniques in SPC focus on detecting anomalies in the parameters being assessed and raise an alert once the parameter goes out of a certain threshold. This concept translates well to biosurveillance and syndromic surveillance where the parameter being assessed is disease incidence and an outbreak is signified by the change in the distribution of disease incidence data. Most of the early event detection algorithms are designed to work in this manner - an alert is raised once the disease incidence crosses a certain threshold. The thresholds need to be set at such a level that they produce meaningful alert. One of the biggest problems with existing systems is that they produce too many alerts, leading public health officials to ignore them. An ignored alert is as good as no alert.

Additionally, early event detection algorithms need to be designed with purpose in mind. In order for the algorithm to be effective at detecting the type of outbreak of interest, it needs to be parameterized with the appropriate data. While the statistical methods behind the algorithms are interchangeable, the algorithms themselves are not. For example, while a certain county might have an early event detection algorithm for influenza outbreaks that can take into consideration seasonality and day of the week effects, in order to apply the algorithm to another county, it will need to be re-parameterized with new data relevant to the new county. Thus, early event detection algorithms are highly dependent on the data available to parameterize them. Some examples of general types of threats that an algorithm can be designed for are;

- Natural outbreak vs. bioterror attack
- Disease specific outbreaks
- Temporal vs. spatial
- General anomaly detector vs. specific event detection

Three SPC statistical methods are commonly used for early event detection algorithms;

1) The Shewhart method and Hotellings X^2 method
2) Cumulative sum statistics (CUSUM) and multivariate CUSUM (MCUSUM)
3) Exponential Weighted Moving Average (EWMA) and multivariate EWMA (MEWMA)

While these statistical methods can be used to monitor any parameter, in the case of biosurveillance they typically look at disease incidence measured either by case counts as a function of time, absenteeism, or sales of some health related product.

The **Shewhart method** is a fairly simple technique that monitors a single data feed (univariate) and compares its distribution to an expected, “normal” distribution. If the observed distribution does not match the expected distribution of data, an alarm is raised. This method is ideal when attempting to detect large shifts in disease incidence but is unable to “remember” previous trends in the data and therefore unable to account for long term changing trends in the baseline incidence of a disease and for smaller shifts in disease incidence. The Shewhart method can be modified to detect smaller shifts in disease incidence by using the “Western Electric” rules which state that if the data comes too close to the threshold (without crossing it) multiple times, to raise an alert. This increased sensitivity comes at the cost of increasing the false alert rate. Typically, the threshold for alert for the Shewhart method is set at three standard deviations from the baseline. Hotelling's X^2 method is a multivariate form of the Shewart method that looks at the multivariate distribution of the data and will raise an alarm if the multivariate data significantly deviates from the expected distribution. Like the Shewhart method, Hotelling's X^2 method can detect large shifts in disease incidence well but is not as good at small shifts and cannot account for long term changes in disease incidence. Additionally, the complexities of addressing the shortcomings of Hotelling's X^2 method are amplified by its multivariate nature. Because multiple data feeds are being used, the false alert rate will be high and determining the appropriate threshold for alert is difficult given that all data streams are not likely of equal utility for indicating an outbreak. Additionally, when an alert is raised, it will require further analysis to determine whether the alert was caused by a rapid change in a single data stream or in multiple data streams.

The cumulative sum statistics (CUSUM) method calculates whether the data observed has changed from one distribution to an alternative distribution. If the cumulative sums of the changes from the expected distribution cross the threshold, an alarm is raised and the sum is reset. Additionally, the data feed, once pre-processed, should be normally or near-normally distributed for it to be used effectively by CUSUM. Because CUSUM tracks the sum of the changes in distribution, it is able to take into consideration historical data, thus giving it a "memory". In contrast to the Shewhart and Hotelling's X^2 methods, CUSUM is less effective at detecting large changes in disease incidence but because of its "memory", it is better at detection of smaller shifts in disease incidence. Multivariate CUSUM shares the same strengths and weaknesses as CUSUM as an early event detection algorithm.

Exponentially weighted moving average (EWMA) and its multivariate form, MEWMA are similar to CUSUM in that they calculate the changes in the distribution of the data but they weight more recent observations as contributing more heavily to triggering an alarm than older observations. EWMA and MEWMA can behave like the Shewhart/Hotelling's X^2 methods or the CUSUM/MCUSUM methods depending on how they are parameterized. Thus, these algorithms can be effective at detecting large or small shifts in disease incidence.

The SPC techniques described above are all *temporal* based methods. They analyze a parameter as a function of time without taking into consideration any spatial elements. *Combining both temporal and spatial elements into analysis is a different kind of integration than described above*. The most commonly used method by the public health community is scan Statistics usually implemented in the **SaTScan software**. This technique counts the number of cases within a geographic circle and varies the time period of observation. The software then randomly varies the radius of the circle and its location for a given region, comparing the observed disease incidence data with the expected. If it detects an abnormal cluster that deviates significantly from the expected number of cases, it raises an alarm.

There are **no** multivariate algorithms currently in operation in any biosurveillance system (Fricker, 2011). Figure 28 describes which univariate statistical methods are being used by some current systems and the parameters of the algorithms.

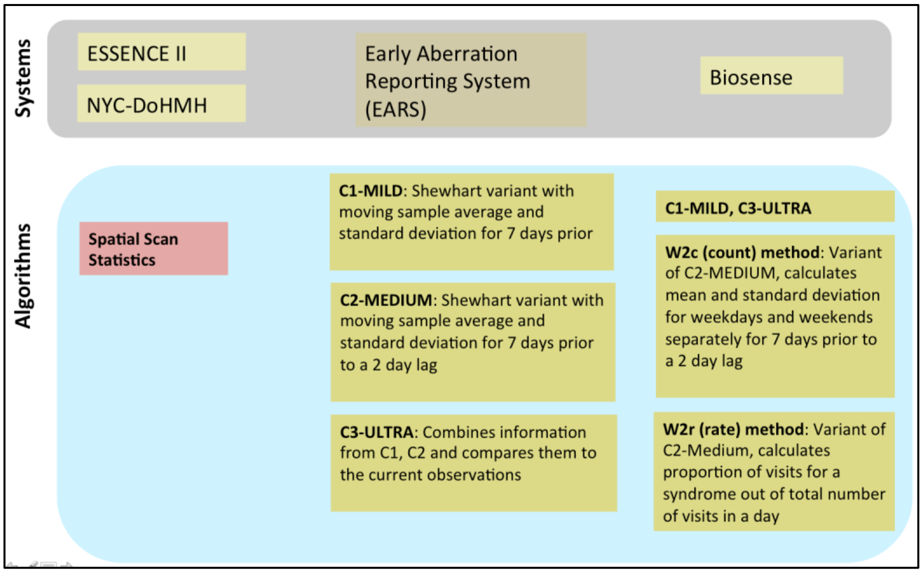


Figure 28: Integration algorithms currently being used by some surveillance systems

To summarize the capabilities of the *multivariate statistical methods* and when they could be applied;

- **Hotelling's X^2 is better at detecting *larger shifts* in disease incidence**
- **MCUSUM is better at detecting *smaller shifts* in disease incidence**
- **MEWMA can be made to *perform similarly* to Hotelling's X^2 and MCUSM depending on how its smoothing parameter, λ, is defined**

Additionally, upon receiving an alert, further analysis would be required due to the multivariate nature of the data. It may be difficult to sort through which of the data feeds triggered the alert since the algorithm looks at changes in the multivariate distribution of the data and not at any single data feed.

There are multiple reasons why most surveillance systems rely on monitoring multiple univariate data streams. Firstly, it is much easier to implement and interpret the information from univariate statistical methods. Even with less sensitivity when compared to multivariate methods, the algorithms being employed for the univariate data stream monitoring already have too high of a false alert rate (Shmueli & Burkom, 2010). The use of multivariate algorithms, at this point, will only likely increase this rate and result in public health officials ignoring the too frequent alerts. Determining how to best set the thresholds for multivariate algorithms to minimize the false alert rate is uncertain. Another issue preventing the adoption of multivariate algorithms is that there is a lack of good methods to validate their efficacy and parameterize them, which both stem in large part from the lack of a "gold" standard data set. While simulated data can be used for parameterization and validation, there are questions as to its value.

Other issues surrounding the data integration and the use of multivariate algorithms are that it is unknown how to use non-syndromic related data with syndromic data. Additionally, one of the assumptions behind the use of SPC methods and, as a result, early event detection is that it assumes that the distributions of outbreak data are constant. However, the distribution of disease incidence is not constant and is likely to change depending on strain and type of pathogen.

It is difficult to evaluate and determine which algorithm is best for early event detection. While some multivariate algorithms have better capabilities for detecting certain kinds of changes in disease incidence, many other factors may prove to be more important when deciding which algorithm to implement. In general, because of the increased sensitivity of multivariate analyses, MCUSUM and MEWMA (if parameterized properly) are more ideal early event detections to use because they can detect smaller shifts in disease incidence, thus taking advantage of the increased sensitivity. However, this increased sensitivity comes at a cost: an increased false alert rate. This excessive false alert rate problem is already an issue for currently operational surveillance systems would only be exacerbated if multivariate algorithms are adopted for analyzing data. Solutions are needed to address this issue before these algorithms can be deployed operationally. One suggestion might be to set the threshold level for alerts, and thus the false alert rate, not at the best level for detection (which tends to result in too many alerts) but to the level that would match the capability of public health departments to respond to and investigate reports. Given that an alert ignored is as useful as no alert, perhaps it may be possible to increase usability of the surveillance systems if the algorithms are designed to alert to the investigational capabilities and resources of the department. Additionally, because of the diverse nature of outbreaks, it is nearly impossible for a single algorithm to be applicable in all scenarios. The algorithms need to be chosen and designed to search for a specific disease or type of outbreak and are dependent on the data available locally. A risk and threat assessment can help elucidate what are the most pertinent types of outbreaks that should be monitored for and algorithms can be designed specifically around the most relevant outbreak risks.

7.0 Progress Summary and Next Steps

Since the last progress report in June 2012, LANL has made significant progress and succeeded in completing the data stream evaluation using two different methods. Table 20 provides a summary of the overall progress and immediate next steps that are proposed for each task. One of the important products that developed following the survey of local, national and international disease surveillance systems was the BRD. We envision the BRD becoming a useful resource for both the national and global biosurveillance community in the years to come and have proposed to upgrade it to a web hosted, free of charge tool.

The evaluation of data stream categories serves to provide information to surveillance system developers, decision makers and public health analysts on what data stream categories could be focused on and offer the best information for the specific biosurveillance goals. In addition, the evaluation framework LANL has developed is a capability that can be used for both absolute and comparative analyses of new data streams as they come on line, or for that matter, be configured for evaluation of surveillance systems, tools used for disease prediction, and/or disease forecasting. It may also be possible to convert the LANL developed evaluation framework into a tool that can be used by analysts working with the DTRA sponsored Biosurveillance Ecosystem (BSVE).

For immediate next steps we propose to vet both the evaluation framework and the values assigned to the data stream categories, with an SME panel composed of members involved in day to day biosurveillance such as public health analysts and practitioners, as well as representatives of the global biosurveillance community. This will help us deliver a more robust evaluation. We also propose to evaluate specific data streams deemed to be a priority by DTRA.

Surveillance window based evaluation used specific data streams, but results have been reported for data stream categories when comparing across the case studies and across methods. However, specific data streams that showed utility for specific diseases have also been identified. A natural transition of this task will be into the development of a surveillance window application/tool that can be integrated into the DTRA sponsored BSVE.

Information presented on integration algorithms that can be used in surveillance systems provides a first look at the landscape of this very much developing field. We also intend to publish at least three manuscripts based on these tasks.

Task	Completed	Product	Immediate next steps
Survey of biosurveillance resources	October 2012	Biosurveillance resource directory	Upgrade to web hosted version Transition to JPEO, Publish paper
Multi Criteria Decision Analysis (MCDA) - based ranking of data stream categories	January 2013	Ranked list of data stream categories	Evaluate with an operational SME panel, Publish paper
Surveillance window based evaluation of data stream categories	January 2013	List of most useful data stream categories	Use case studies for development of surveillance window application, Publish paper
Integration algorithm analysis	January 2013	Report	None

Table 20: Summary of progress

8.0 References

Azar K. Multiattribute decision-making: Use of three scoring methods to compare the performance of imaging techniques for breast cancer. *http://repository.upenn.edu/cis/*. University of Pennsylvania, 2000

Batini C., Cappiello C., Francalanci C., and Maurino A. Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41(3) Article 16 52 pages. 2009. DOI=10.1145/1541880.1541883 http://doi.acm.org/10.1145/1541880.1541883

Bravata DM., McDonald K.M., Smith WM., Rydzak C., Szeto H., Buckeridge D.L., Haberland C., Owens DK. Systematic Review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases. Annals of Internal Medicine. 140(11):910-922, 2004.

Buehler JW. Review of the 2003 National Syndromic Surveillance Conference - Lessons learned and questions to be answered. MMWR Weekly. 53(Suppl):18-22, 2004.

Buehler JW., Sonricker A., Paladini M., Soper P. and Mostashari F. Syndromic surveillance practice in the United States: Findings from a survey of state, territorial, and selected local health departments. Advances in Disease Surveillance. 6(3):1-20, 2008.

Centers for Disease Control and Prevention. Epidemic/epizootic West Nile virus in the United States: guidelines for surveillance, prevention, and control. Division of Vector-Borne Infectious Diseases, 2003.

Corley CD., Lancaster MJ., Brigantic RT., Chung JS., Walters RA., Arthur RR., Bruckner-Lea CJ., Calapristi A., Dowling G., Hartley DM., Kennedy S., Kircher A., Klucking S., Lee EK., McKenzie T., Nelson NP., Olsen J., Pancerella C., Quitugua TN., Reed JT. and Thomas, CS. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science. March 2012, 10(1): 131-141. doi:10.1089/bsp.2011.0096.

Fraser C., Donnelly CA., Cauchemez S., Hanage WP., Van Kerkhove MD., Hollingsworth TD., Griffin J., Baggaley RF., Jenkins HE., Lyons EJ., Jombart T., Hinsley WR., Grassly NC., Balloux F., Ghani AC., Ferguson NM., Rambaut A., Pybus OG., Lopez-Gatell H., Alpuche-Aranda CM., Chapela LB., Zavala EP., Guevara DE., Checchi F., Garcia E., Hugonnet S., Roth C. and the WHO Rapid Pandemic Assessment Collaboration. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. Science. 324(5934): 1557-1561 2009. DOI: 10.1126/science.1176062

Fricker, RD. Some methodological issues in biosurveillance. Statist. Med., 30: 403–415. 2011. doi: 10.1002/sim.3880

Hitchcock P., Chamberlain A., Wagoner MV., Inglesby TV. and Tara O'Toole. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science. 5(3): 206-227. 2007. doi:10.1089/bsp.2007.0041.

Hubalek Z and Halouzka J. West Nile Fever – a reemerging mosquito-borne viral disease in Europe. Emerging Infectious Disease. **5**:643, 1999.

Iowa State University, “West Nile Virus Infection”, The Center for Food Security and Public Health, 2009.

Komar N *et al.* Experimental infection of North American birds with the New York 1999 strain of West Nile Virus. Emerging Infectious Diseases. 9:311, 2003a.

Komar N. West Nile Virus: epidemiology and ecology in North America. Advances in Virus Research. 61:185, 2003b.

Molaei G et al. Host feeding patterns of Culex mosquitoes and West Nile Virus transmission, northeastern United States. Emerging Infectious Diseases 12:468, 2006.

Pipino LL., Lee YW. and Wang, RY. Data quality assessment. Commun. ACM. 45(4):211-218, 2002. DOI=10.1145/505248.506010 http://doi.acm.org/10.1145/505248.506010

Shmueli G and Burkom H. Statistical challenges facing early outbreak detection in biosurveillance. Technometrics 52 (1), 2010.

Trpis M and Hausermann W. Dispersal and other population parameters of Aedes aegypti in an African village and their possible significance in epidemiology of vector-borne diseases. American Journal of Tropical Medicine and Hygiene 35:1263, 1986.

Wallenius http://www.sal.tkk.fi/vanhat_sivut/Opinnot/Mat-2.4194/TDDA/Dyer_July2007rev5.pdf, 2007.

WHO http://www.who.int/topics/social_environment/en/

Wonham MJ., de-Camino-Beck T. and Lewis MA. An epidemiological model for West Nile virus: invasion analysis and control applications. Proceedings of the Royal Society B 271: 501, 2004.

Younger M., Morrow-Almeida HR., Vindigni SM., Dannenberg AL. The built environment, climate change, and health: Opportunities for co-benefits. American Journal of Preventive Medicine. 35(5): 517-526, 2008.

9.0 Appendix A: BRD Development

Over the past two years, the BRD has undergone significant evolution from a static Excel spreadsheet to a dynamic, searchable, relational database that contains 296 records as of October 2012. Figure 29 shows this evolution to the most current desktop version of the database. In its current form, the platform for the BRD is Filemaker Pro and data can be exported from it in various formats. A plan has been developed for long term maintenance, curation, updating and global access for this resource.

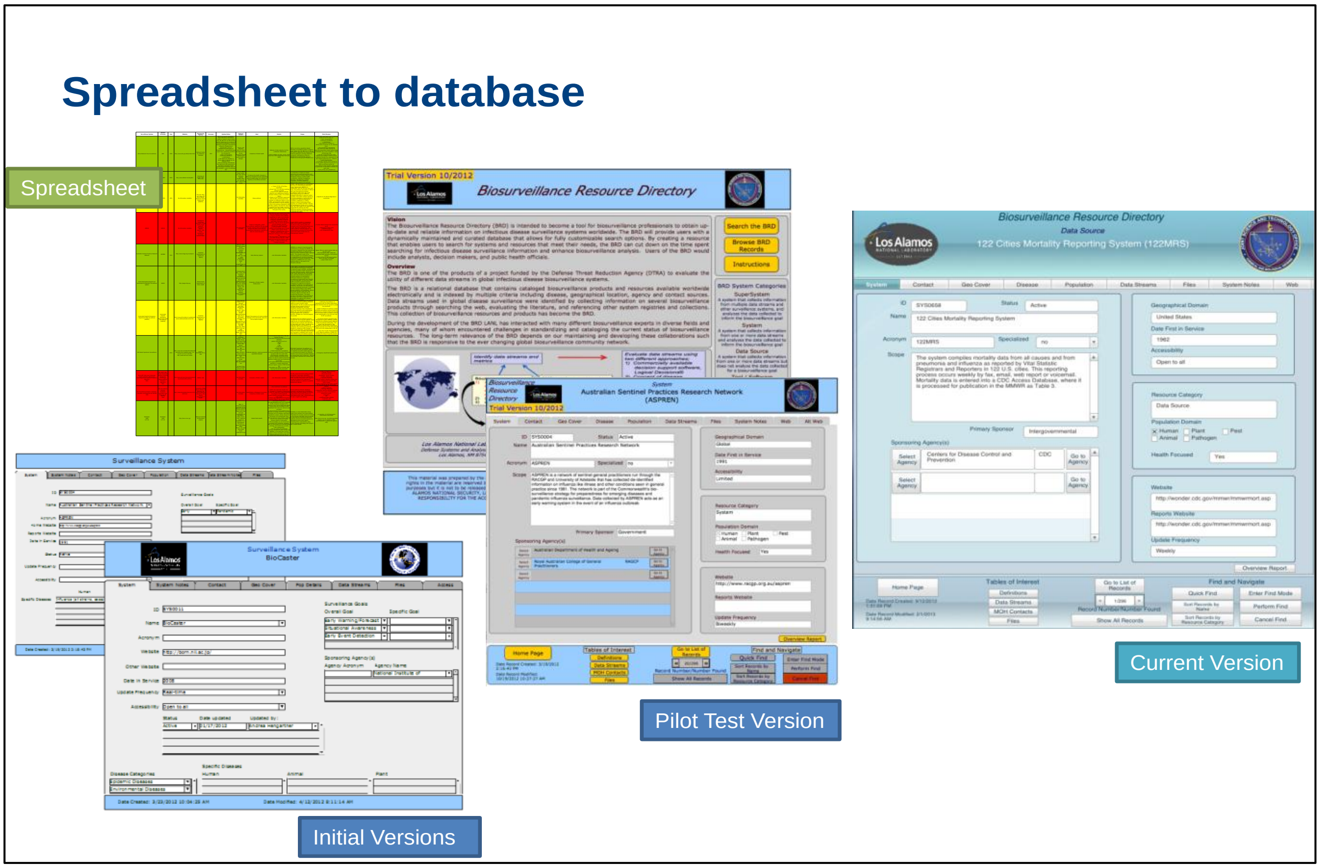
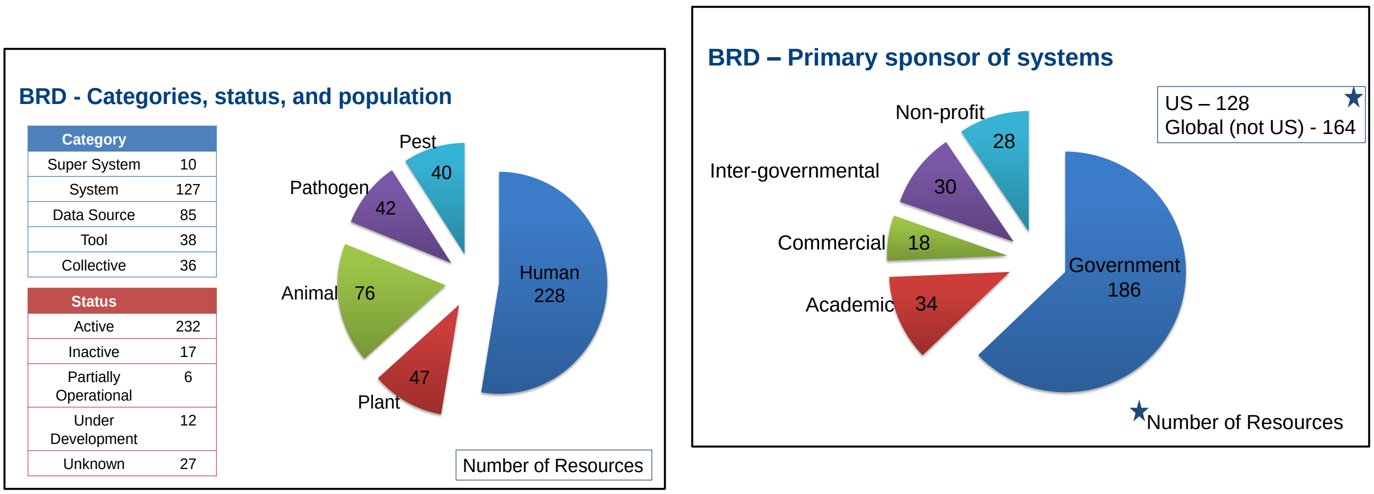


Figure 29: Evolution of data collection of biosurveillance resources

Figure 30a shows the information captured in the BRD in terms of the categories of information, the status of those records and the distribution of the records between human, animal and plants. Evaluators of the BRD expressed interest in maintaining records that were inactive to provide historical perspective and offer information for developing efforts. Not surprisingly, a majority of the records have a human focus. In terms of the distribution of sponsors for the resources captured in the BRD, a majority of them are funded by federal (US and global) and state governments (Figure 30b). Captured in the information collected about the system (if available) is the name and acronym of the system, the date the system was in service, the accessibility of the system (is it open to all, or are there limitations to access), the primary sponsors of the system, if the system is associated with GIS functionality, and if the system's primary surveillance focus is health. Also collected is contact information, information regarding the scope and domain of the system, the diseases of interest pertinent to that the system, and the geographic and population coverage of the system. Websites associated with the system are directly accessible from the database. Data stream information is also captured based on our developed data stream framework. If the system uses other specified systems for data gathering or analysis, then that is also captured on the data stream tab, and is directly linked within the database to those systems. Figure 31 shows the distribution of data stream categories used by the local, national and international surveillance systems recorded in the BRD, and interestingly, the data stream categories that have shown the most utility through our evaluation methods seem to be the ones primarily used.



Figures 30a and b: Content and sponsor statistics for the BRD

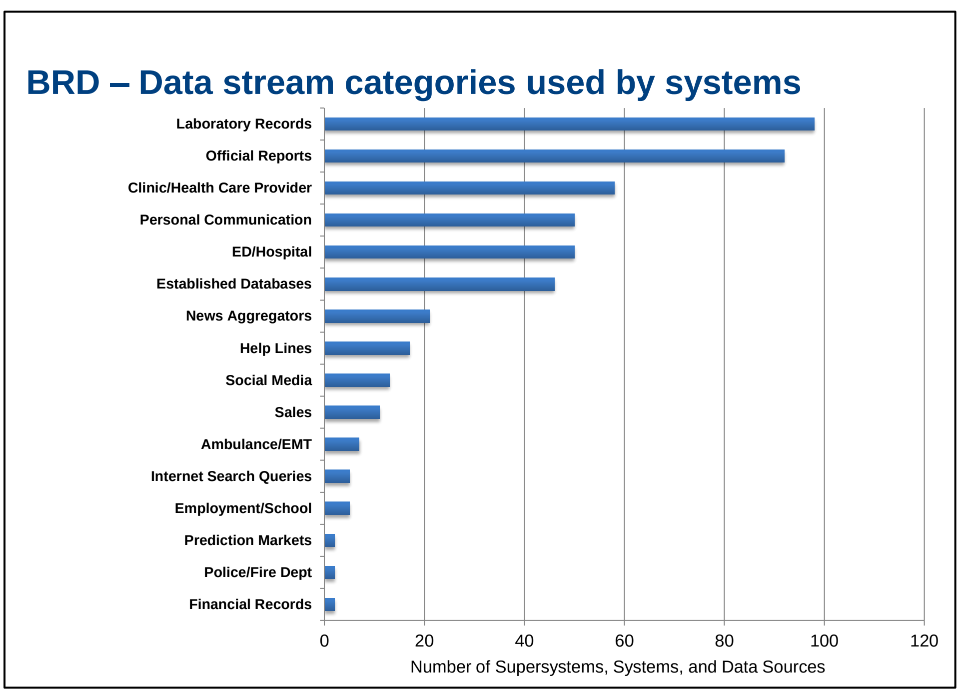


Figure 31: Data stream statistics for the BRD

10.0 Appendix B: BRD pilot test results

The BRD underwent pilot testing both formally and informally and assessments were reviewed and used in the development of a plan to sustain, curate and make the BRD globally accessible. The formal pilot testing of the BRD was performed by 14 experts in biosurveillance that represented 13 agencies including those with expertise/interest in plant, animal, and human infectious disease biosurveillance. Table 21 lists the names and affiliations of these experts.

Aaron Scott Director, National Surveillance Unit	**USDA / APHIS / VS** Animal and Plant Health Inspection Service Veterinary Services
Judy Akkina Biological Scientist, Epidemiologist	**USDA / APHIS / VS** Animal and Plant Health Inspection Service Veterinary Services
Sarah Tomlinson Associate Coordinator, NAHLN	**USDA / APHIS / VS / NAHLN** Animal and Plant Health Inspection Service Veterinary Services, National Animal Health Laboratory Network
Tracy McNamara Professor of Pathology	**Western University of Health Sciences, Veterinary School**
Russ Bulluck Associate Director, Emergency Response Center for Plant Health Science and Technology	**USDA / APHIS / PPQ** Animal and Plant Health Inspection Service Plant Protection and Quarantine
Amy Kircher Associate Director, NCFPD	**NCFPD / UMN** National Center for Food Protection and Defense, University of Minnesota
Luther Lindler Science Advisor	**AFHSC** Armed Forces Health Surveillance Center
Michael Latham Senior Public Health Analyst Associate Director of Policy	**CDC** Public Health Surveillance and informatics Program Office (PHSIPO) Office of Surveillance, Epidemiology and Laboratory Servicies (OSELS)
Jim Writer / Matthew Bachtell Biological Scientist/Information Specialist	**NCMI** National Center for Medical Intelligence
Courtney Corley Health Security and Informatics Research Scientist	**PNNL** Pacific Northwest National Laboratory
Sheri Lewis Global Disease Surveillance Program Manager	**JHU/APL** The Johns Hopkins University Applied Physics Laboratory
Laurel Boyd Epidemiologist	**State of Oregon/OHA** Oregon Health Authority
Michael Colleta Lead Informatics Analyst	**NACCHO** National Association of County and City Health Officials
Charles Ishikawa Associate Director, Public Health Programs	**ISDS** International Society for Disease Surveillance
Amy Kircher Associate Director, NCFPD	**NCFPD / UMN** National Center for Food Protection and Defense, University of Minnesota

Table 21: BRD Pilot Testers

A fillable survey was made available to the pilot testers and included a set of questions organized into a general assessment, an assessment of the content of the BRD and an assessment of the functionality. Appendix C is the survey questionnaire that was sent to the evaluators, Appendix D includes feedback received by each evaluator for these questions. Several use cases were identified by testers;

- As a reference directory
- As a central site for all available surveillance programs in a given region
- To identify biosurveillance systems/sources that have specific characteristics
- To identify new resources that can be used for a variety of biosurveillance goals
- To understand local, regional, and global biosurveillance
- To find potential contacts
- To network with other jurisdictions and learn about new approaches or best practices in surveillance

Table 22 summarizes the feedback received following a review of the content of the BRD. Feedback was obtained for all categories of content and the value for a majority of the categories of information was deemed to be high, expect for document links, reports and population coverage, where opinions varied considerably.

Content	Value	Comments
Contacts	High	MOH highlighted, Networking and relationship building important
Agencies	High	Links helpful
Scope, Status, Funding	High	Accuracy and credibility must be assured
Supersystems Systems Data Sources Tools Collectives	High	All resources important, Expand to include local and regional capabilities, lab surveillance systems, and more foreign data streams Also include groups that are working on maintaining or implementing data stream standards
Content	**Value**	**Comments**
Data Streams	High	Relevant information- would like even more detail, and links or information about data access
Documentation Links	High/Low	Reports and reference material especially helpful for those not associated with health research libraries
Reports	High/Low	"Essential", "Not Needed"
Geographical Coverage	High	Increase granularity, show catchment area of system/ data stream
Population Coverage	High/Low	Expand Include demographics
Disease Coverage	High	Expand detail Expand beyond infectious disease

Table 22: BRD content review

A summary of feedback obtained for improving the functionality for the BRD is shown in Figure 32 along with the use frequency. Again, the opinions varied widely between evaluators, although everyone agreed that the use frequency would increase were the BRD to be maintained and regularly curated and updated.

BRD evaluation: functionality and use estimates

Current	Suggested
Desk-top install	Web link / host
User interface	Update, make more intuitive
Search and find	Update, faceted, allow for custom searching, and drill-down
Results as list	Put results on map
Direct links to websites	Keep web-linking

- **As is:** Routine, As Needed, Daily, Monthly, Occasionally
- **If updated and curated regularly:** Frequency generally would increase (i.e. from monthly to weekly)
- **Suggested update frequency:** As frequently as possible, weekly, monthly, semi-monthly, annually

Figure 32: BRD feedback for functionality

An informal evaluation of the BRD was conducted at the annual conference organized by the International Society for Disease Surveillance in December 2012 where the BRD was made available for interested attendees to use and test. In addition, a poster describing the BRD was also on display. The overall feedback was very positive and there was much interest in when and how the BRD could be made accessible to users. Some of the entities represented by viewers of the poster and the BRD included Black and Veatch, the University of Illinois at Urbana-Champaign National Center for Supercomputing Applications, the Department of Health for California, EWIDS, Redwood MedNet, DHS OHA Division of Health Threats Resilience, NCB-Prepared, USDA Veterinary Services, Oregon Health Authority and ISDS.

11.0 Appendix C: Biosurveillance Resource Directory (BRD): Test User Survey

Name (optional):

Job Title (optional):

Organization:

Please briefly describe your job responsibilities and field of work:

Please spend some time browsing and searching the BRD. After you have familiarized yourself with the BRD we would appreciate your answers to the following questions.

Overview Questions

1. Please describe how you would use the BRD in your field of work.

2. Please tell us about the aspects of the BRD that you found *most* useful

3. Please tell us about the aspects of the BRD that you found *least* useful

4. What was your overall impression of the BRD?

5. For whom do you think the BRD would be most useful?

6. Would you recommend the BRD to others?

7. How often would you use the BRD, in its current version, if the BRD were available to you?

    - Daily
    - Weekly
    - Monthly
    - Occasionally
    - Not very often
    - Never

8. How often would you use the BRD if the BRD were dynamically curated and updated?

    - Daily
    - Weekly
    - Monthly
    - Occasionally
    - Not very often
    - Never

Utility Questions

9. Which search features did you find useful?

10. Were you able to search for the information you wanted?

- yes
- no

If no, please describe why.

11. What was your impression of the overview report? What would you add or take away?

12. Did you find the links to external web pages helpful?

- Yes
- No

13. Which web links did you use?

- Resource (System, Data Source, Tool)
- Agency
- Contact
- File

14. How important to you is access to the reports/journal articles?

Content Questions

15. Were the Systems and Supersystems comprehensively represented in the BRD?

- Yes
- Mostly
- No - I know of many systems that are not represented in the BRD
- Don't know

16. Were there any important systems /supersystems that were not included?

17. Which statement best describes the content collected about each system/supersystem?

- Nearly all the systems records adequately described the status and scope of the biosurveillance system
- Many of the systems records adequately described the status and scope of the biosurveillance system
- Not very many of the systems records adequately described the status and scope of the biosurveillance system
- Hardly any of the systems records adequately described the status and scope of the biosurveillance system

18. Would you include any additional information about systems/supersystems in the BRD? If yes, what would you include?

19. Did the inclusion of Collectives, Data Sources and Tools enhance or detract from the utility of the BRD?

20. What areas would be most beneficial to you for the BRD to expand? (Mark all that apply)

- Systems/Supersystems data records
- Collective data records
- Data Sources data records
- Tools/Software data records
- Contacts
- Diseases
- Geographical Coverage

Other - please describe

21. How often do you think the BRD should be actively updated? (adding new resources, changing status/contact/scope information of current systems, etc.)

- Daily
- Weekly
- Monthly
- Semi-annually
- Annually

22. Other Comments

12.0 Appendix D: User Survey Responses

BRD Feedback (freeform answers - surveys not filled out);

Charlie Ishikawa, ISDS

Mike Latham, CDC

Sheri Lewis, JHU/APL

Russ Bulluck, USDA-APHIS-PPQ

Charlie Ishikawa, ISDS

Feedback received by email after reviewing desk-top version

Content: System categories to an software call out process explicitly as an collect process or analyze collective isn't the best name return SME panel missing state local public authorities territorial etc. academics researchers and developers Content Testing geographic area of subject Population coverage demographic description of subjects Disease categories disease and outcome categories these don't seem internally consistent Journal article would be great to link to articles and other resources Love the web site interface

Content is great. Not sure the categories align with the use case I envision local, state or government epidemiologists or health officials taking to the tool.

Use Case that I have in mind: User approaches BRD as they are applying or asking for public/government resources to implement a public health intervention or prevention program. As part of the program, the user needs to develop an evaluation that will gauge or measure whether or not the intervention is achieving its objectives (e.g., decrease incidence of disease X, or decrease stress on healthcare system due to Y behavior). For the evaluation, the user needs data to develop their evaluation.

Content Thoughts: Overall, the content is there to partially meet the use case that I had in mind. Especially useful are the links to articles, websites and contacts. If there were additional information of additional/alternative ways to categorize the records, it would better support the use case. Below are a few ideas:

- Tools: Data processing (e.g., natural language processing) is a functionality of interest
- The SME Panel: Missing public health authorities
- Every state and territory should have a reportable disease or conditions system. Would be helpful to have a link to statutes.
- Every state should also have some kind of health alert action network.
- Additional information about the systems and data streams:
    - Catchment area (geographic)?
    - Demographic characteristics of the subjects?
    - Representativeness (e.g., proportion of ED visits monitored by the system)?
- Disease categories: Might be better organized or sub-categorized per causative agent and mode of transmission…at least.
- Accessibility: Accessible to whom? For example, government authorities users, government officials or employees, general public with authorization, etc.
- It might be helpful to also include, under collectives, groups that are working on developing or maintaining data stream standards (e.g., S&I Framework).

Utility Thoughts: The interface should be enhanced to increase the accessibility of the content and hence the tool's utility. Here are the observations I made:

- The search functionality could be more intuitive or similar to search utilities that are more commonly used (e.g., Google)
- Current search function is cumbersome, or at a least I wasn't able to master it within 60 minutes. It would be good to have a free-text method for searching the records across all fields.
- On my Mac (see profile) the interface did not present on one display; I needed to use two displays and could still not move it all on to one screen
- Consider grouping the tools by critical tasks in the ISDS Final Recommendation: Core Business Processes and EHR

Requirements for Public Health Syndromic Surveillance. Those groupings are an operational view of surveillance systems in general and one the resonates with practitioners.

- Love the links to websites and the ability to display the site within the window.

Mike Latham and team, CDC

Feedback received by conference call after reviewing desk-top version

“Great product!” “Thumbs up from everyone in the room”

Roadmap and Biosurveillance implementation plan has as key component landscape/catalog of what capabilities are currently available

CDC registry - no longer funded, will not remain active but will be archived

Content: how great the fidelity should be, how much utility - Increase fidelity in regards to local level

Find out regionalized efforts in collaborative surveillance

Useful for predictive analysis / forecasting

Interested in future collaboration:
BSV working group
Federal Registry group
CDC registry group

Curation: possible a charter or agreement among stake-holders
Users and resource owners are stake-holders, promote value added
Manual updates will be required

Keep CDC in the loop

Overview report essential

Web-hosting essential

Comments on LANL hosting through research library:
Premier research library - access to state-of-the-art search and find
Student population - high school to post-doc for curation would keep costs down

Would like to “put their data in the BRD “ to “get it out there”

Sheri Lewis, JHU/APL

Feedback received while reviewing desk-top version at ISDS

Important resource, necessary
Difficulty with downloading/installing desk-top version - need to webhost
How will the BRD be curated and kept up-to-date?

Russ Bulluck

Feedback received be email after reviewing desk-top version at ISDS

Not bad. A number of useful plant-related resources. Several, however, were limited due to login/password protection (esp for foreign govts) Also, numerous links to nucleotde sequences were of limited value, but good to know that they are there if needed.
25 or so links to plant sources, so not too bad. Some things I had not seen, many of which are from PPQ (at least 5), but that’s good too.

13.0 Appendix E: SME Panel Survey, Summary Report

13.1 Introduction

LANL has recently been funded by the Defense Threat Reduction Agency (DTRA) - Joint Science and Technology Office (JSTO) for a project to determine the relevance of data streams and data integration schemes for an integrated global biosurveillance system, focused on human, plant, and animal infectious diseases. A key requirement for this evaluation is the development of defined metrics and use of defined methodology. LANL is using two different methods for evaluating data streams. First, a decision support tool called Logical Decisions® that assigns utility scores to data streams based on weighted metrics and data stream assigned values for the metrics. Second, a concept called surveillance window developed at LANL that assigns a window of time specific to a disease within which information coming from various data streams can be determined to have utility. The disease specific surveillance windows are also dependent on operations and surveillance goals. LANL is using the following approaches for defining metrics and identifying traditional and non-traditional data streams:

1. Perform a comprehensive survey of current and planned surveillance systems that cover human, plant and animal diseases and operate locally, nationally or globally.
2. Establish a subject matter expert (SME) panel and survey the panel to obtain information about data streams and metrics.
3. Conduct an extensive review of the scientific literature pertaining to the field of biosurveillance.

This report describes LANL's effort to establish an SME panel and summarizes the results of survey responses. In addition to requests for information about data streams and metrics for evaluation, LANL's survey questions included general questions about various terms used in the field of biosurveillance, opinions about strategies for effective biosurveillance and gaps seen by SMEs.

13.2 SME Panel and Participation

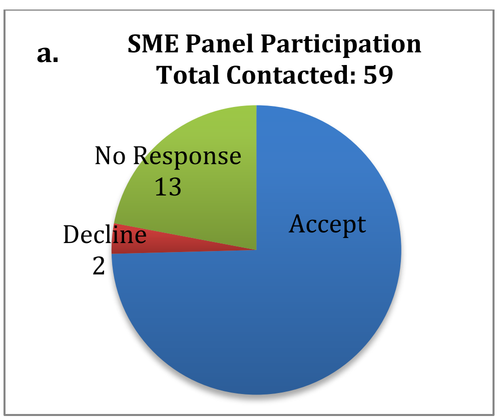
A list of potential SMEs was contacted based on the individual's experience with biosurveillance and his/her affiliation with current biosurveillance applications and needs. While the primary purpose of the panel was to get information about data streams and metrics, we also decided that it would be helpful to understand the context in which each SME understood biosurveillance, and therefore created a questionnaire that asked both broad and specific questions related to global biosurveillance and data stream utility.

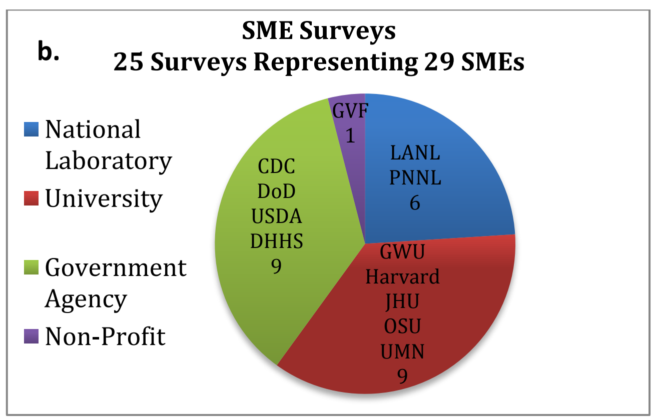
Over the course of 6 months panel acceptance and survey information was received. This report is a summary of responses obtained from the unclassified SME panel. Following is a list of institutions/entities that participated in the unclassified SME survey:

Institutions Represented by Survey Responders
AFHSC Armed Forces Health Surveillance Center
CDC Centers for Disease Control and Protection
GVF Global Viral Forecasting, Inc.
GWU The George Washington University
Harvard University

JHU-APL Johns Hopkins University - Applied Physics Laboratory
LANL Los Alamos National Laboratory
OASD, NCB/CB Office of the Assistant to the Secretary of Defense for Chemical and Biological Defense
OSU Oklahoma State University
PNNL Pacific Northwest National Laboratory
DHHS, USPHS Department of Health and Human Services, United States Public Health Service
UMN University of Minnesota
USDA, APHIS, CPHST Animal and Plant Health Inspection Service, Center for Plant Health Science and Technology
USDA, NAHLN National Animal Health Laboratory Network
USDA, **NSU** National Surveillance Unit

Figure E1 provides statistics on the total number of participants, level of participation and representation of various entities. It is important to note, that biosurveillance experts outside of the US were not contacted at the time of writing this report, although efforts are ongoing to obtain more information from SMEs at the World Health Organization (WHO), the World Organization for Animal Health (OIE), and the Food and Agriculture Organization of the United Nations (FAO). We contacted a total of 59 SMEs, of those contacted 44 were willing to be panelists. However, we were only able to obtain 25 survey responses representing 29 panelists (due to three surveys representing joint responses: CDC, USDA-NSU, and AFHSC). These responses were most valuable and contributed significantly to both the identification of broad data stream categories as well as development of defined metrics.





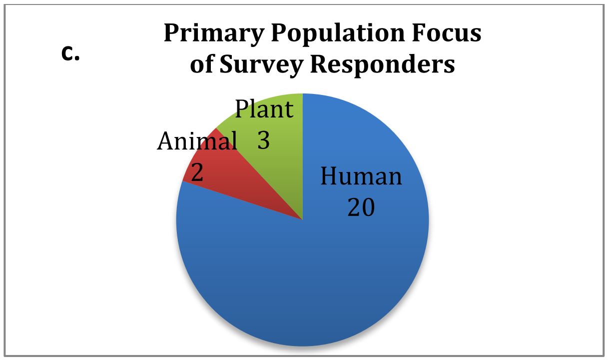


Figure E1: SME Panel participation and formation**. a.** Total number of SMEs willing to be on panel. **b.** Number and primary organization of SME panelists completing the survey. **c.** Population focus of survey responders.

11.3 Survey Responses

The following sections describe the results obtained from the panel survey. Sections are primarily organized by the specific questions that were asked to the SMEs, although we found that some questions and responses fit better when reorganized by topic. A summary of the answers is provided and, where possible, statistics on responses have been included to facilitate understanding of a majority opinion.

I. BIOSURVEILLANCE DEFINITIONS AND GOALS

Survey Q: What is *your* brief definition of the following terms: biosurveillance, *global* biosurveillance, and *integrated* biosurveillance?

Biosurveillance: A definition of biosurveillance as described by the SMEs usually encompassed at least two parts: The first was the goal of biosurveillance such as situational awareness, early detection of a health event or early warning of an event. The second was the means by which the goal would be attained such as targeted monitoring of populations or collection and analysis of health related data. A third part to the definition was added by two SMEs and included outcomes desired, such as mitigating adverse health effects or initiating control programs. Similarly, Homeland Security Presidential Directive 21 (HSPD 21) defines biosurveillance as:

> "the process of active data-gathering with appropriate analysis and interpretation of biosphere data that might relate to disease activity and threats to human or animal health —whether infectious, toxic, metabolic, or otherwise, and regardless of intentional or natural origin—in order to achieve early warning of health threats, early detection of health events, and overall situational awareness of disease activity."

Although this definition was cited by 5 of 24 SMEs, the majority of SMEs had a narrower focus: surveillance and detection *of disease outbreaks caused by infectious pathogens*, rather than the broader scope defined by HSPD21. **Global biosurveillance** definitions were generally similar but with a global/worldwide perspective and covering human, animal and plant diseases.

Integrated Biosurveillance: While the definition of biosurveillance for most SMEs concentrated on human health and disease surveillance, the definition of integrated biosurveillance expanded the biosurveillance definition to include many more diverse data sources and populations. Additionally most SMES included the caveat that the diversity was useful only if the data streams could be effectively combined for meaningful analysis. Two of the best definitions, given by experts in animal disease surveillance, are shown below.

The following definition of comprehensive biosurveillance was given to describe global biosurveillance, but it we feel it is a very good summary of integrated biosurveillance:

> "Comprehensive surveillance describes a surveillance system at many levels within and across diseases, species, and systems. Comprehensive surveillance includes diverse types of health indicators, relies on various data sources, and includes all aspects of the surveillance process. Comprehensive surveillance follows a specific plan, is objective driven, is coordinated at all levels, and is standardized, allowing for multiple-level (including national) conclusions to be made."

Additionally the same team of experts provided this definition for integrated surveillance:

> "The combination of surveillance system components that have common characteristics to increase efficiency. Integration can occur on many levels; for example, multi-disciplinary and harmonized planning, implementation, and analysis of surveillance systems across disease and species; standardized performance metrics to allow comparable information from different surveillance systems; and testing the same sample for multiple diseases (only when epidemiology and surveillance objectives justify that the cost efficiency gained does not detract from the quality and utility of surveillance information received). The information management component of a surveillance system particularly benefits from integration in allowing for an efficient exchange of information among the various stakeholders and across different surveillance systems, resulting in a system that allows for entry of and ready access to multi-source information and optimal resource allocations by objective to avoid duplication."

Survey Q: What, *in your expert opinion,* are the primary goals of global biosurveillance?

As the goals of biosurveillance were almost universally included in the given definitions, responses to the survey question regarding biosurveillance goals are also included here. LANL has created a framework to broadly categorize biosurveillance goals. Based on our analysis (again primarily through consultation and a thorough literature review) four broad goals were identified: early warning of health threats, early detection of health events, situational awareness, and consequence management. LANL's definitions of surveillance goals are the following:

Early Warning of Health Threats: Surveillance that enables identification of potential threats including emerging and re-emerging diseases that may be undefined or unexpected.

Early Detection of Health Events: Surveillance that enables detection of disease, outbreaks (either natural or intentional in origin), or events that have occurred, but are not yet identified.

Situational Awareness: Surveillance that monitors the location, magnitude, and spread of an outbreak or event.

Consequence Management: Surveillance that assesses impacts and determines response to an outbreak or an event

Baseline Awareness: Information that can inform and facilitate the achievement of the above surveillance goals and can be related to population demographics and health, the natural, political, and social environment, and underlying disease patterns and characteristics.

The goals tend to follow a time-course from early warning to consequence management, although there is certainly overlap in time. Underlying all of the goals is the need to have baseline awareness of disease and environmental determinants. The SME panel listed many goals of biosurveillance. Frequently mentioned were early warning, early detection and situational awareness either broadly or with greater specificity (such as preventing disease spread for early warning). Additionally, SMEs considered the difference in goals from a military (force health protection) perspective versus a universal public health perspective.

Figure E2 shows the goals mentioned by the SME panel as they fit into LANL's framework. The broad goals have been linked and overlapped to indicate that there is no absolute cut off on a time scale when any one surveillance goal would be deemed irrelevant. Likewise, Baseline Awareness is a significant requirement to achieve any of the surveillance goals identified in the figure.

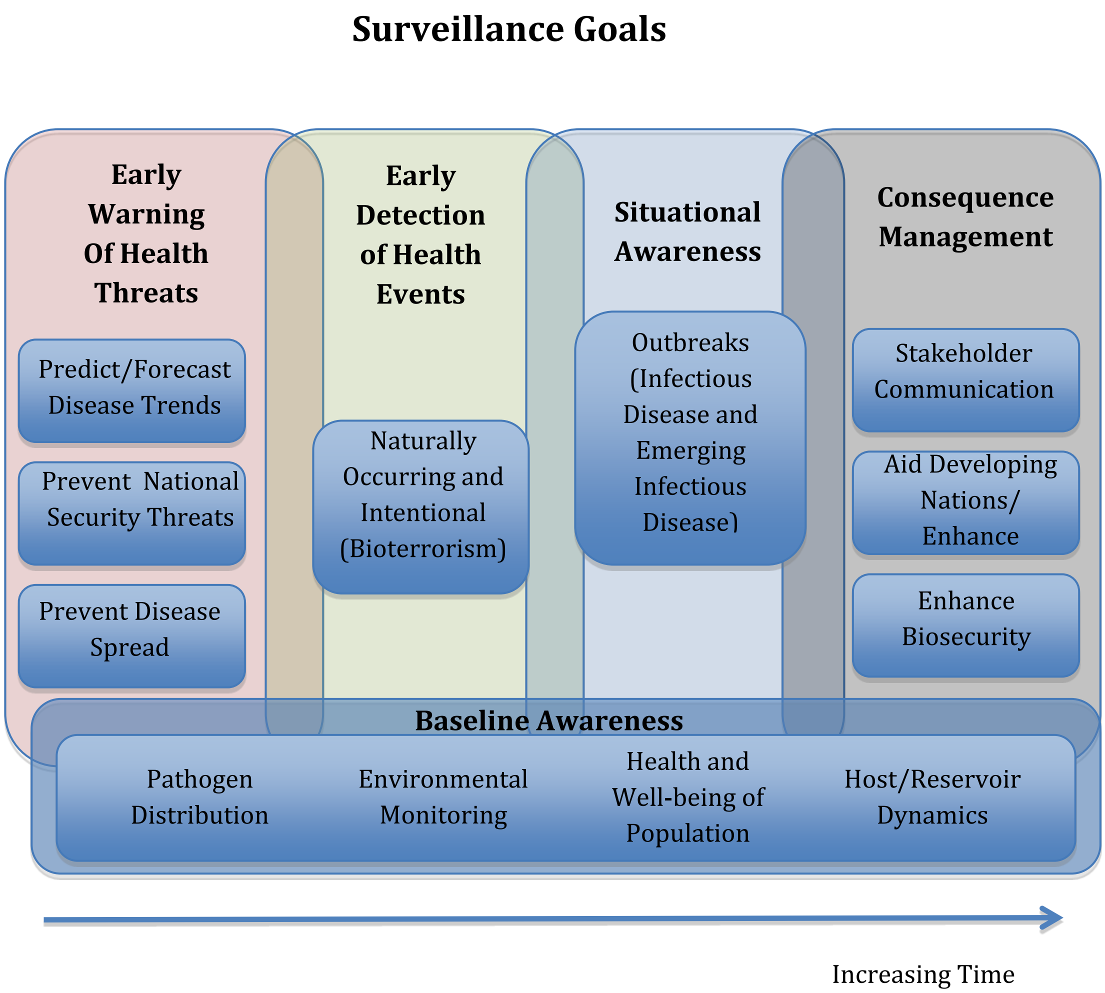


Figure E2: Biourveillance goals

Survey Q: Do you think a single integrated global biosurveillance system can fulfill all goals of surveillance?

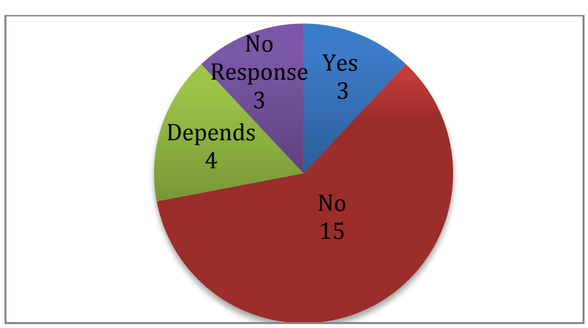


Figure E3: Response of SME panel to the question, "Do you think a single integrated global biosurveillance system can fulfill all goals of surveillance?"

A majority of SMEs do not believe a single system can fulfill all biosurveillance goals (Figure A3). The primary reasons given for this were:

- The system would be too big with too many goals
- The system would not be flexible or detailed enough
- It would be politically difficult to partner effectively with all stakeholders (countries, agencies, the private sector…)
- The benefits of multiple systems to provide checks and balances would be lost
- "Pie in the Sky"

Those SMEs that thought an integrated system might be possible qualified the response by indicating that to be successful a "bottom up" approach should be taken and that the political issues associated with a global biosurveillance system would have to be surmounted.

II. TRADITIONAL AND NON-TRADITIONAL DATA STREAMS

Survey Q: What is *your* brief definition of the following terms: data stream, data stream integration, and non-traditional data stream?

Data Stream: Most SMEs considered a data stream to be a single source of information or data that could be used in a biosurveillance system. A majority of SMEs indicated a preference that the data stream should be continually accessible on a regular basis (real-time or near real time, daily, monthly).). It was also recommended that the term "data stream" be replaced by "data source" as "stream" may imply a continuous feed of data, when many data sources did not actually provide this.

Based on SME responses, and the diversity of what was considered important, LANL has collated the responses into a definition that could be considered a definition of an *ideal* data stream:

> "A single source of unique, timely (real-time), and spatially relevant information that is standardized and collected in a quantity and class that is needed for meaningful results, that targets a specific population, that is available at many scales (from molecular to ecosystem), is electronically available in both raw and reportable form, and has been rigorously validated. "

Data Stream Integration: The overwhelming consensus among SMEs was that data stream integration combines multiple sources of information. Several SMEs considered data stream integration for specific purposes such as looking for trends in pathogen distribution, or for enhancing predictive ability and modeling analysis. Additionally, data stream integration was considered a means for increasing the spatial and temporal scope of surveillance as well as including plants and animals in surveillance. Several SMEs also highlighted data structure requirements for integration such as the need for common file formats or data fusion necessary for leveraging the most information and analysis from integration.

The following definition, (also provided by the experts in animal disease surveillance) sums up much of the above:

> Data Stream Integration - "the meaningful fusion of disparate data streams into biosurveillance models or systems that increase the specificity, sensitivity, positive predictive value, and negative predictive value of predicting, detecting, or forecasting an infectious disease more than using any one data stream independently."

Non-Traditional Data Stream: The request for this definition is directly related to LANL's project and was included to add to the list of data streams we obtained through our surveillance system survey and literature review. Most SMEs considered non-traditional data streams as sources of information that are new or innovative, have not been traditionally applied or used for a specific biosurveillance purpose, or are streams that are not health or disease focused. Figure E4 lists the various non-traditional data streams identified by the SMEs.

While it is interesting to note the wide variety of streams considered non-traditional by the SME panel, the distinction between traditional and non-traditional seems to be blurred. For example, the data streams identified by the SMEs were only one component of the process in determining LANL's final list of data stream categories that we are subsequently using in our data stream analysis.

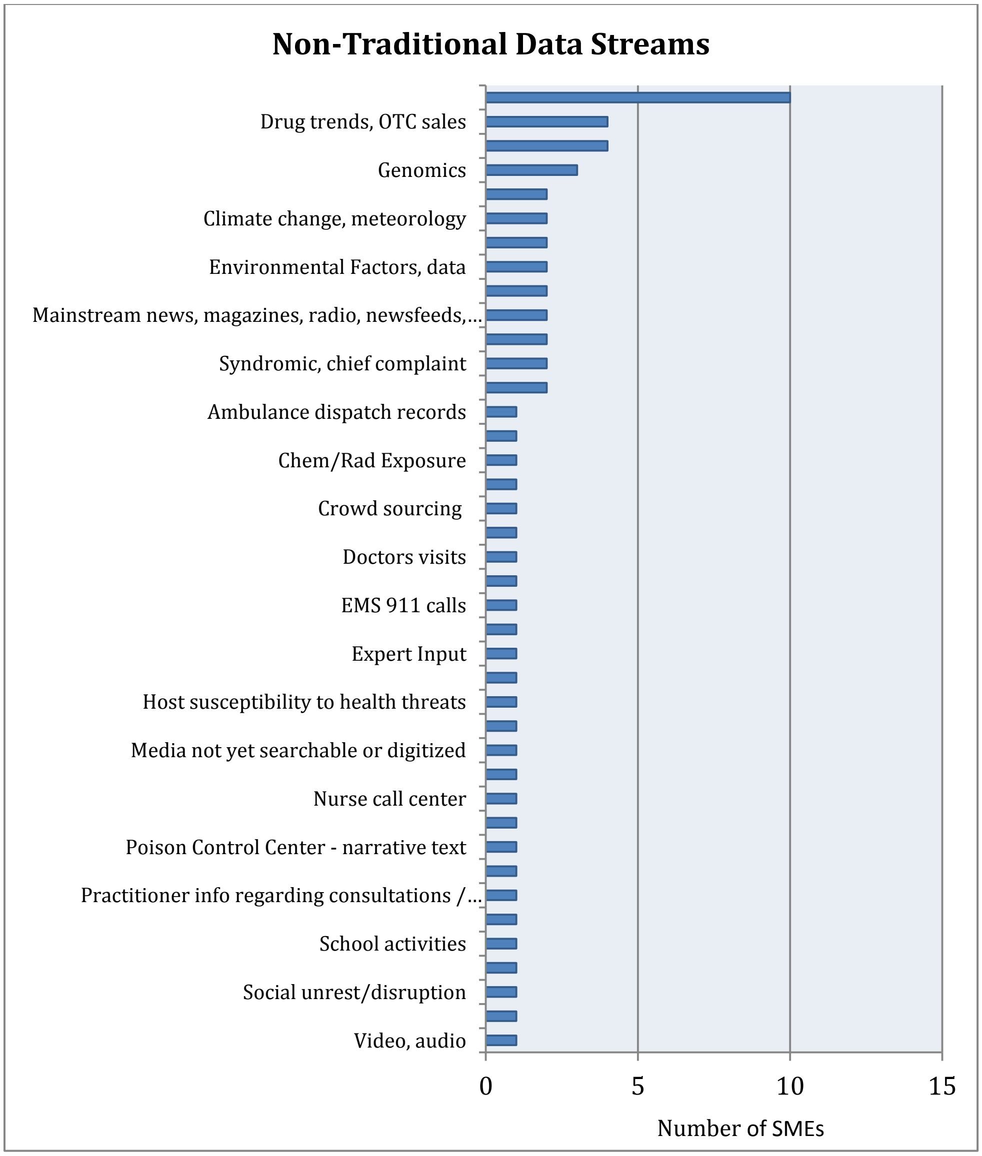


Figure E4: Non-traditional data streams as identified by the SME panel.

III. DATA STREAM METRICS

Survey Q: How would you evaluate the utility of a data stream to be used in global biosurveillance? Can you identify a set of metrics?

This question directly relates to LANL's evaluation of data streams. Many metrics were mentioned by the SMEs, and a diverse set of terms was used to describe similar metrics. Therefore, in order to evaluate the metrics, we placed the SME metrics into broad categories. We determined these categories based on our literature review, our surveillance systems collection, and on this survey. Table E1, column 1 shows LANL's categories of metrics and column 2 lists the metrics provided by the SMEs sorted into what we determined to be the appropriate category.

Several SMEs mentioned usefulness or utility as a metric, and it is included in our statistics, however, this metric is not included in the table below since the usefulness or utility of a data stream is what we are ultimately evaluating.

LANL's Metric / Definition	SME Metrics Fit to LANL's Metric
Accessibility The extent to which the data stream is available	Feasibility of obtaining appropriate data, availability, ease of accessibility, ease of collection, portability, difficulty of use, ability to query
Cost The cost to set-up, operate, and maintain the data stream	Cost of data collection, cost
Credibility The extent to which the data stream is considered reliable and accurate	Acceptability, consistency, reliability, credibility of source, trustworthy, quality, completeness of data, comprehensiveness, existence and inherent sensitivity of screening tests, data associated with lab support, data associated with positive lab results, involvement of WHO, follow IHR guidelines, applicability to IHR, robustness
Flexibility The data stream's ability to be used for more than one purpose (such as for use in surveillance for more than one disease, or for more than one goal)	Flexibility, diseases covered
Integrability How well the data stream can be integrated or linked/combined with other data streams	Transferability, standardizable, interlinkable, integrable, feasibility of implementation
Geographic /Population Coverage The geographic or population area of coverage	Geographic coverage, population coverage
Granularity The level of detail of the data stream	Detail, granularity, representativeness of data
Specificity of Detection The ability of the data stream to identify the event, pathogen, or disease outbreak of interest	Probability of disease detection, ability to signal and detect, actionable information, indicators and warning, specificity, relevance (to disease of interest), accuracy of detection
Sustainability The data stream's continued availability over time	Sustained accessibility, longevity, sustainability
Time to Indication The time required for the data stream to indicate detection of a disease, outbreak, or event	Time to disease detection, time to incident detection (compared to traditional sources), latency, identification (time to disease), performance, sensitivity
Timeliness Earliest time that the data is available	Data available in real/time or near real time, timeliness of reporting, how often data updated, timeliness

Table E1: Column one lists the set of metrics and the associated definitions determined by LANL. Column two lists the metrics provided by the SME panel sorted according to LANL's metrics.

Survey Q: Can you rank the metrics in order of importance?

Several SMEs made the point that the ranking of metrics would differ based on surveillance scenarios and data type. For those SMES willing to rank metrics the top metrics listed by the panel (using the above broad categories) are listed in Figure E5. We used three different methods to evaluate rank. The first was to determine the most frequently mentioned metric and the top five metrics most commonly mentioned are in the first column of Figure E5. Second, we counted the metrics that were most commonly mentioned as ranking in the top 4, and these are listed in the middle column of Figure E5. The last method was to count only the top ranked metric from each SME, and the results are shown in the third column. The metrics that were consistently ranked, regardless of the method of evaluation, were Time to Detection, Credibility, and Specificity of Detection. Even though Cost was the third most frequently mentioned metric, it did not rank high enough to be even in the top four overall ranked metrics.

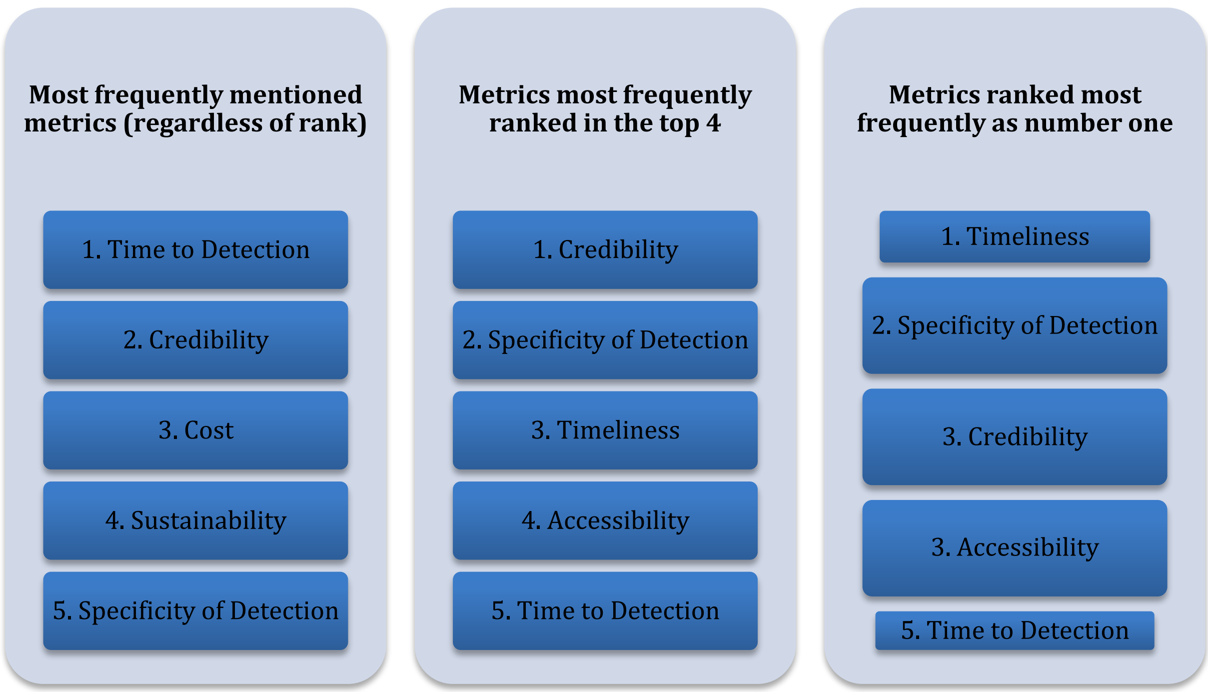


Figure E5: SME ranking of metrics by Column 1) Most frequently mentioned, Column 2) Most frequently ranked in the top four, and Column 3) Most frequently ranked as number one.

IV. DISEASES

Survey Q: What in your opinion would be the 10 most important diseases that we could use to evaluate data streams for biosurveillance?

This question directly relates to our second approach for data stream evaluation that uses disease specific "surveillance windows" to evaluate the data streams. LANL needed to get a list of priority diseases to develop outbreak simulations or timelines based on historical outbreaks.

The top diseases mentioned by the SMEs spanned a wide range and included diseases developed from bioterrorist pathogens, WHO reportable diseases, emerging diseases, and common flu. Animal and plant diseases that are listed below were mentioned almost exclusively by SMEs whose area of expertise is in either plant or animal surveillance. Interestingly, FMD and Rinderpest were two animal diseases that crossed disciplinary boundaries. One suggestion for picking priority diseases was to pick those that represented the most commonly occurring syndromes – febrile, respiratory, gastrointestinal, antimicrobial resistant, and sexually transmitted.

Bacterial: Anthrax, Brucellosis, Plague, Q-fever also known as Balkan grippe, Tularemia, Shiga toxin - producing E. Coli, Salmonellosis, Shigellosis, Botulism, Tuberculosis, TB, Cholera, Lyme disease, Tick-borne Encephalitis (TBE), Community-acquired MRSA

Viral: Smallpox, Monkey Pox, Ebola, Marburg, Lassa Fever, Bolivian hemorrhagic fever, Machupo, Crimean Congo HF, Rift Valley Fever, Hemorrhagic fever with renal syndrome (HFRS), Hantavirus pulmonary syndrome (HPS), Japanese, Encephalitis, Nipah virus, Hendra virus, encephalitis, West Nile encephalitis, Yellow Fever, Dengue, DHF (global), Chikungunya Fever, Avian Influenza/, Highly pathogenic avian influenza, SARS Severe Acute Respiratory Syndrome, Measles, Mumps, HIV/AIDS, Influenza (all strains, seasonal and pandemic), Noroviruses, Norwalk-like viruses, RSV Respiratory Syncytial Virus infections , Eastern Equine encephalitis, Venezuelan Equine encephalitis,

Prion, Fungal, Protozoan: BSE (cows)/ new variant Creutzfeldt-Jakob disease (humans), Malaria, Coccidioidomycosis- valley fever, lung disease, Leishmaniasis, Visceral Leishmaniasis, (VL)

Non-Specific: Unknown emerging, Asthma exacerbations, Pneumonia, Hemorrhagic Fever, Gastro-intestinal, Conjunctival, Food borne, Emerging/unknown, National Notifiable Diseases, Bioterrorism diseases

Animal Diseases: FMD, Rinderpest, Exotic Newcastle disease, African Swine Fever, Classical Swine Fever

Plant Diseases: Plant pathogens and associated vector distribution, Wheat stem rust, US strain, Wheat stem rust, strain Ug99 (Middle East), Rice blast, Soybean rust, Soybean/corn rust (Agro disease), Cassava blight, Pierce's disease of grapevines, Citrus greening (Huanglongbing), Sudden oak death

The SME survey was combined with other lists obtained from national and international agencies (CDC, DoD, WHO) to determine potential diseases to use in our evaluation. We narrowed down this list by choosing 10-15 diseases that, as a group, met the following criteria:

- Are representative of different modes of transmission
- Are representative of different host species (human, animal, plant)
- Can be used in an evaluation for all data streams
- Detailed data can be found about each disease and associated outbreaks

V. SYSTEMS AND TECHNOLOGY

Survey Q: What gaps do you see in current biosurveillance systems/strategies?

The panel was in general agreement that many significant gaps exist, especially in the context of an integrated global biosurveillance system. Answers focused on addressing problems in four broad areas:

- People/policy
- Data
- Resources
- Systems

Political/policy challenges, data integration needs, and data access/data format gaps were most commonly mentioned. Additionally a lack of trained specialists in multiple areas of biosurveillance (public health, analysts, modelers) and a lack of capabilities in low-resource settings/ limited funding were also frequently mentioned. The chart below (Figure E6) shows the identified gaps and the percentage of SMEs indicating the gap.

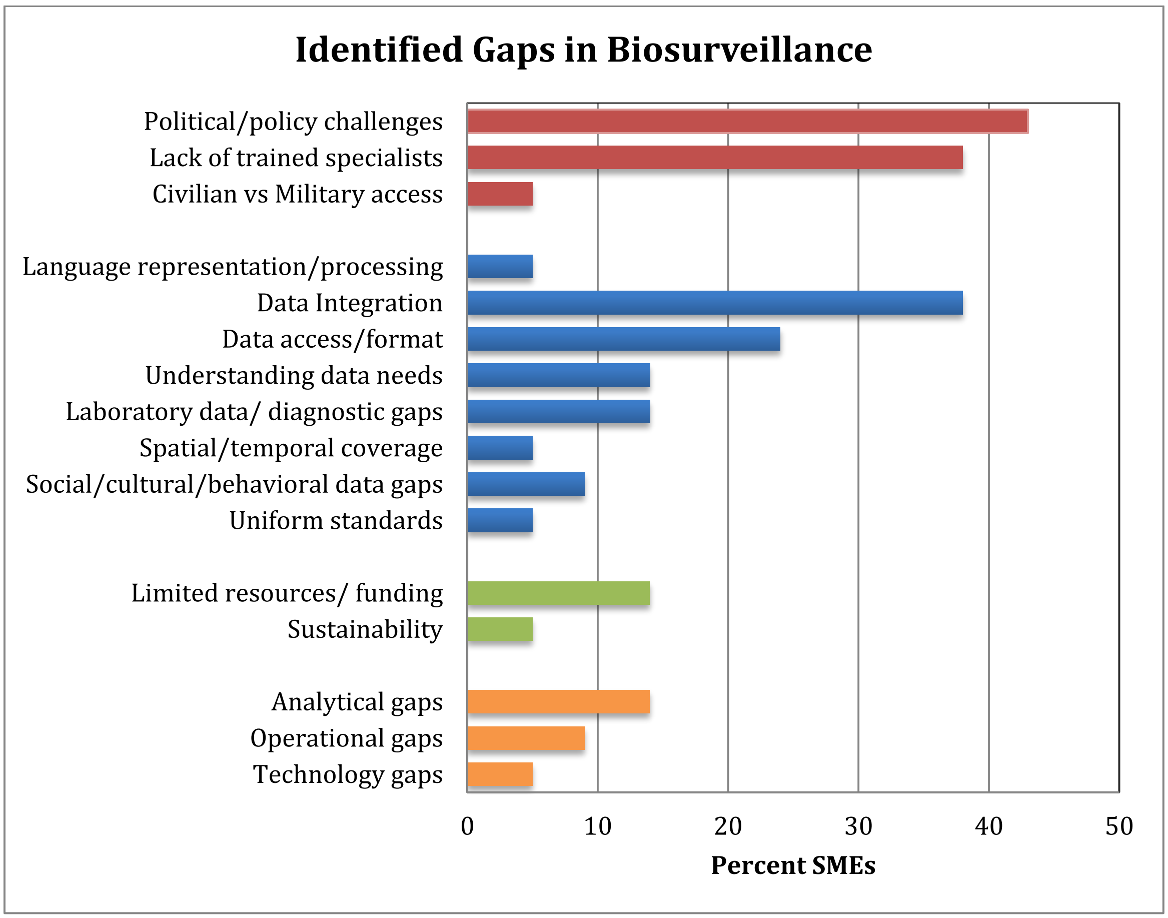


Figure E6: Current gaps in biosurveillance systems and strategies as identified by the SME panel.

Survey Q: What *current* technologies do you think are most important to a global biosurveillance system? What *near-future* technologies do you think will have greatest utility to a global biosurveillance system?

A listing of the panel's response to both of these questions is found in Table E2. In the context of near future technologies, no completely new technology was identified. Rather, many responses indicated a need to improve upon technologies currently available: make a rapid lab test faster, convert data to electronic form, or increase use of already available technology such as using cell phones for disease tracking and reporting. There was also no one technology that stood-out as being a magic bullet for a global biosurveillance system. Instead it is clear that the technological challenges will continue to be broad and interdisciplinary.

Current Technologies (Number SMEs)	Near Future Technologies (Number SMEs)
IT platform for data integration (5)	IT platform indicating data stream utility/impact (2)
Statistical process control, scan statistics, Algorithms (3)	Integrated data systems (across laboratory networks) (4)
Bioinformatics and modeling of disease occurrence and spread (3)	Bioinformatics, computational biology, machine learning, pattern recognition, data mining (2)
Micro-array diagnostics High throughput genomics analysis Sequencing (3)	Multiplex Diagnostic technology, Molecular diagnostics and microfluidics (3)
Detection systems Disease diagnostic systems (4)	Pathogen screening system and other diagnostics in low resource environments (4)
Handheld devices (3)	Cell phone technology for disease tracking and reporting (3)
Cellular phones (3)	Wireless technology (2)
Social media (3)	Social media (2)
Internet (2)	Cloud computing (1)
Electronic Medical record implementation (1)	Electronic Medical Records (4)
Current/Active Disease surveillance systems (2)	Surveillance system portals (1)
Environmental data streams (land-sat, demographics, airline and maritime) (1)	New sensor technologies (2)
The Analysts (1)	Historical adverse events registry to provide info on patterns and indicate trends (4)

Web-based language translators (1)	Web-based language translators (1)
	Health insurance claims (1)
	Predication markets (1)
	Improved point-of-care diagnostics, enhanced diagnostics (1)

Table E2: Current and near-future technologies identified by the SME panel as having the most importance or greatest impact. The number of SMEs that mentioned the technology is listed in parentheses.

13.4 SUMMARY

We are grateful to those willing to participate on our SME panel and respond to our survey questions. Answers provided were thoughtful and from multiple perspectives that allowed us to put together a compilation of information that should be useful to the biosurveillance community and is a step towards unifying the language of biosurveillance. Our goal is to facilitate the building of surveillance systems using a common framework by which disparate systems, data streams, and tools can be evaluated in order to develop more robust surveillance systems.

14.0 Appendix F: Methods of Determining Values of Metrics

Outlined below are the methods of how the values for the inputs for the metrics were determined. As each reviewer was going through the relevant literature of the data stream category they were evaluating, they would determine the input values into the metrics based upon the performance of the data stream using the criteria outlined below. Additionally, the relationships between these values and *utility*, the common unit that LDW uses, are described here. Two types of measurements for the values could be used: quantitative values and labels. Quantitative Values are numbers that reflect the metric. Labels are text descriptions that describe the level of performance on a metric.

1. **Accessibility**

Definition: The extent to which the data stream is available.

Accessibility is measured as a label with three options: Difficult Accessibility, Medium Accessibility, and Easy Accessibility.

- ***Difficult Accessibility***- is when the data stream being analyzed has been used in at least ones system and faces many (3 or more) obstacles in data access
- ***Medium Accessibility***- is when the data stream being analyzed has been used in at least one system and faces some (less than 3) obstacles in data access
- ***Easy Accessibility***- is when the data from a particular data stream is freely accessible

Utility increases as accessibility becomes easier.
Examples of obstacles include: privacy concern, passwords, subscription, membership/ group affiliation, non-digitized information, etc.

2. **Cost**

Definition: Cost is defined as the cost to set-up, operate, and maintain the data stream

- ***High Cost***—there is a cost to obtain, set-up, and maintain the data stream
- ***Medium Cost***—there is a cost for only two of three of the following: to obtain, set-up, or maintain the data stream
- ***Low Cost***—there is a cost for only one of the following: to obtain, set-up, or maintain the data stream

The *utility* decreases as cost increases.

3. **Credibility**

Definition: The extent to which the data stream is considered reliable and accurate

Credibility is measured as a label with three options: Low Credibility, Medium Credibility, and High Credibility.

- ***Low Credibility***- is when the data stream being analyzed provides limited actionable results. Additionally, data must be validated by another source

- ***Medium Credibility***- is when the data stream being analyzed provides actionable results but data still requires validation
- ***High Credibility***- is when the data stream being analyzed provides actionable results and minimal validation needed.

The *utility* increases with higher credibility.

An actionable result refers to when the data provided by the data stream is of high enough quality that it can be acted on. Validation refers to the need to confirm the data from the data stream using a separate source.

4. Flexibility

Definition: The data stream's ability to be used for more than one purpose (such as for use in surveillance for more than one disease, or for more than one goal)

Flexibility is measured as a label with three options: High Flexibility, Medium Flexibility, and Low Flexibility.

- ***High Flexibility***- is when the data stream being analyzed can be used for more than three purposes
- ***Medium Flexibility***- is when the data stream being analyzed can be used for two purposes
- ***Low Flexibility***- is when the data stream being analyzed can be used for only one purpose

The *utility* increases with more purposes the data stream can be used for.

Examples of purposes include: diseases, events, goals, types of surveillance etc.

5. Integrability

Definition: How well the data stream can be linked/combined with other data streams

Integrability is measured as a label with four options: Extremely Integrable, Highly Integrable, Moderately Integrable, and Not Very Integrable.

- ***Extremely Integrable***- is when the data from the data stream is in a structured and standardized format and has been integrated with one other data stream in more than one biosurveillance system
- ***Highly Integrable***- is when the data from the data stream is structured and in a standardized format and has been integrated with one other data stream in one other biosurveillance system
- ***Moderately Integrable***- is when the data from the data stream is either in a structured format or has been integrated with one other type of data
- ***Not Very Integrable***- is when the data from the data stream is unstructured and has never been integrated with another type of data.

The *utility* increases if the data stream is in a standardized format and if it has been integrated with other types of data.

Structured data implies xml or other electronic format. A standardized format implies ICD-9, or other agreed upon reporting formats. Integrated refers to the data from one data stream being combined or linked to the data in another data stream.

6. **Geographic/Population coverage**

Definition: The geographic or population area of coverage

Geographic/ Population coverage is a label with four options: Local, Regional, National, and Global.

The *utility* increases as the Geographic and Population coverage becomes broader (i.e. more global)

7. **Granularity**

Definition: The level of detail of the data stream

Granularity is measured as a label with four options: Individual, Community, Regional, and National.

- ***Individual-*** the data is applicable at the individual level
- ***Community***- the data is applicable at the community level
- ***Regional***- the data is applicable at the regional level
- ***National***- the data is applicable at the national level

The *utility* increases as the unit of data tracked becomes smaller.

An individual level refers to a person. A community level is anything from a household to a metropolitan area of a large city. A regional level refers to state or a state-like entity or a grouping of states.

8. **Specificity of Detection**

Definition: The ability of the data stream to identify an outbreak, event, disease, or pathogen of interest

Specificity of Detection is measured as a label with four options: High Specificity, Medium Specificity, Low Specificity, and Indirect Specificity.

- ***High Specificity***- the method of detection for the data stream is disease specific
- ***Medium Specificity***- the method of detection for the data stream is disease category specific (e.g. viral, bacterial, etc.)
- ***Low Specificity***- the method of detection for the data streams tells you syndrome-based information (e.g. ILI, etc.)
- ***Indirect Specificity***- the method of detection for the data stream is an indirect indicator of disease
-

The *utility* is highest for High Specificity and lowest for Indirect Specificity.

9. Sustainability

Definition: The data stream's continued availability over time

Sustainability is measured as a label with two options: Yes and No.

Yes- the data stream is still in use / existence
No- the data stream is not still in use / existence

The *utility* is higher for data streams that are still in use / existence.

10. Timeliness

Definition: Earliest time that the data is available

Timeliness is measured as a label with four options: Near Real Time, Quick, Intermediate, and Slow.

- ***Near Real Time***- The data is available within one day
- ***Fast***- The data is available between one day and one week
- ***Intermediate***- The data is available between one week and one month
- ***Slow***- The data is available after one month

The *utility* decreases as the longer it takes for the data to become available.

11. Time to Indication

Definition: The time required for the data stream to first signal a disease, outbreak, or event

Time to Indication is measured as a label with four options: No Indication, Near Real Time Indication, Medium Indication, and Long Indication.

- ***Long Indication***- The data stream indicates an event, situation or disease outbreak after one week
- ***Medium Indication***- The data stream indicates an event, situation or disease outbreak between one day and one week
- ***Near Real Time Indication***- The data stream indicates an event, situation or disease outbreak within a day.
- ***Indirect Indication***- The data stream does not directly indicate an event, situation, or disease outbreak

The *utility* decreases the longer it takes for the data stream to indicate an event, situation or disease outbreak.

Table 23 depicts the utility score assigned to each metric value. The utility function was set to be the default, linear.

Metric	Label	Utility Score
Accessibility	Easy	1
	Medium	0.5
	Diificult	0
Cost	High	0
	Medium	0.5
	Low	1
Credibility	Low	0
	Medium	0.5
	High	1
Flexibility	High	1
	Medium	0.5
	Low	0
Geo/Pop	Global	1
	National	0.667
	Regional	0.333
	Local	0
Granularity	Individual	1
	Community	0.667
	Regional	0.333
	National	0
Integrability	Extremley	1
	Highly	0.667
	Moderately	0.333
	Not Very	0
Specificity of Detection	High	1
	Medium	0.667
	Low	0.333
	Indirect	0
Sustainability	Yes	1
	No	0
Time to Indication	Long	0.333
	Medium	0.667
	Near Real Time	1
	Indirect	0
Timeliness	Slow	0
	Intermediate	0.333
	Fast	0.667
	Near Real Time	1

Table 23: Metric utility scores

An electronic survey was generated online and distributed through Qualtrics.com. It consisted of six questions. When asked to rank the metrics by biosurveillance goal, the participants of the survey had to drag and drop the metric by order of importance to them. The metrics when presented to them on the survey were randomized. Definitions were provided for each biosurveillance goal and metric.

Participants were not allowed to rank metrics as being equal in importance. Name and affiliation of the participants were asked. Note, when the survey was conducted the metric “Time to Indication” was referred to as “Time to Detection”, thus on the survey Time to Detection is the metric that is being ranked. Table 24 lists the responders for the survey.

The final ranked lists were based on the average rank for the metric over all the responses and rounded to the nearest integer. The rationale for the rounding is that the difference between values at the tenth and hundredth decimal place are insignificant. Metrics whose average ranks were identical were considered to be of equal preference and were indicated by giving the same rank.

Name	Affiliation
Kirsten McCabe	LANL
Alina Deshpande	LANL
Jennifer Foster Harris	LANL
Lauren Castro	D-3, LANL
Elizabeth Hong-Geller	B Division LANL
Sara Del Valle	LANL
Kristen Margevicius	LANL
Mac Brown	LANL
Kristin Omberg	LANL
Larissa May	GWU
Michael Schmoyer	HHS
Russ Bulluck	USDA
Richard Stouder	ORNL
Jeanne Fair	LANL
Prachi	CDC
Richard L. Stouder	ORNL
Dr. Annette Sobel	University of Missouri
Cos DiMaggio	The Tauri Group
Unknown	Unknown

Table 24: Responders to the survey for metric weights